DOCUMENT RESUME

ED 187 717 TM 800 192

AUTHOR Borich, Gary D.
TITLE A State of the Art Assessment of Educational Evaluation.
PUB DATE [80]
NOTE 102p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Decision Making; *Educational History; Educational Philosophy; Elementary Secondary Education; *Evaluation Methods; *Models; Needs Assessment; *Program Evaluation; Research Methodology; State of the Art Reviews; Systems Approach; Values
IDENTIFIERS Dewey (John); Evaluation Problems; Naturalistic Research

ABSTRACT

Some of the historical developments which have influenced educational evaluation are the importance of operationalism and behavioral objectives; curriculum reform; the Elementary and Secondary Education Act; and school and teacher accountability. Major problems related to current evaluation practices include negative attitudes toward evaluation, and the lack of theory and adequate guidelines for evaluation. Traditional conceptual schemes of evaluation include measurement as evaluation; determination of congruence; professional judgment; and applied research, though most evaluators and researchers implicitly support a separation between these two modes of inquiry. Current approaches to evaluation and evaluation models include Provus' discrepancy evaluation model, Stake's antecedents-transactions-outcomes model, and the context, input, process, product (CIPP) model. These decision-oriented evaluation models can be contrasted with value-oriented evaluation, based on Dewey's conceptualization of valuation. Out of this theoretical base, naturalistic evaluation methods have arisen, including the responsive, judicial, transactional, connoisseurship, and illumination models. Concurrently, systems-oriented evaluation has also been developed, to interconnect the planning, development, and evaluation process. All these approaches have implications for needs assessment and policy analysis. (Over 80 bibliographical references are appended.) (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187717

TM800 192

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. Wilder

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A State of the Art Assessment of Educational Evaluation

Gary D. Borich[1]

The University of Texas at Austin

I. Overview of Evaluation

One purpose of educational evaluation is to provide decision makers with information about the effectiveness of an educational program, product, or procedure. Within this perspective, evaluation is viewed as a process in which data are obtained, analyzed, and synthesized into relevant information for decision making.

While most evaluation activities fit comfortably within the bounds of this definition, the specific approach used and procedures employed vary from one evaluation study to another as a function of who is doing the evaluation, the context in which the evaluation is to occur and the desires and needs of the individual or agency contracting the evaluation. While there is basic agreement about the fundamental role of evaluation in education, beyond this there is considerable variance in the conceptual frameworks used by practitioners. Indeed, even the ways in which evaluation has been defined in the literature has produced considerable debate.

Bloom, Hastings and Madaus (1971) point to five different facets of evaluation, not all of which are included in other definitions. These authors pose a broad view of evaluation consisting of the following activities:

1. Acquiring and processing the evidence needed to improve the students learning and the teaching.

---

[1]The author would like to thank Ron Jemelka for his many significant contributions to this paper, especially those pertaining to the concept of value-oriented evaluation which will appear in Jemelka, R. and G. Borich, Traditional and Emerging Definitions of Educational Evaluation, Evaluation Quarterly, in press.

MAR 14 1980

2. Employing a great variety of evidence beyond the final paper and pencil examination.
3. Clarifying the significant goals and objectives of education and determining the extent to which the students are developing in these desired ways.
4. Instituting a system of quality control in which it may be determined at each step in the teaching-learning process whether the process is effective or not and if not, what changes must be made to insure effectiveness.
5. And, ascertaining whether alternative procedures are equally effective or not in achieving a set of educational ends (p. 7-8).

As general as these activities may appear, they are not the only purposes for which evaluations can be conducted. Stufflebeam et al. (1971), for example, divide evaluation into a four part process consisting of context, input, process and product evaluations, each with its own objectives and methods, while Provus (1971), Stake (1967), Hammond (1973), Metfessel and Michael (1967) as well as others conceptualize and partition the process, if not the domain, of evaluation in still other ways.

With evaluators differing on such basic issues, it is not surprising that one can find numerous evaluation paradigms or "models" in the literature to help shape and guide evaluation activities. The problem for the evaluator becomes one of choosing the conceptualization or model most appropriate to his evaluation problem. Because the evaluation models appearing in the literature are purposely general so as to be applicable to a wide variety of educational problems, the task of choosing that conceptualization of evaluation most

appropriate to a specific purpose becomes even more arduous. One focus of this paper will be to trace the origins of the problem of choosing the correct conceptualization or model for an evaluation and to identify some of the underlying factors which have contributed to the heterogeneity of opinion concerning the definition, nature and scope of educational evaluation.

To this end I will present an overview of some historical developments which have influenced the growth of educational evaluation. This chronology will provide the foundation for an interpretation of contemporary movements in the field and the extrapolation of these movements to the not-to-distant future.

Before proceding a personal note is in order. I have struggled in this writing to keep separate the idea of where the field of evaluation is going from the idea of where this author believes it should be going. As most authors will attest any writing is inextricably tied to the author's background, training and philosophy and this chapter is no exception. As Kuhn (1970) has made us painfully aware "an apparently arbitrary element, compounded of personal and historical accident, is always a formative ingredient of the beliefs espoused by a given scientific community (and scientist) at a given time....among those legitimate possibilities, the particular conclusions he does arrive at are probably determined by his prior experience in other fields, by accidents of his investigation, and by his own individual makeup" (p. 4). Kuhn's observation leads us to ask who might be the wiser: the scientist who writes about his field influenced by his own implicit biases and the philosophy of his scientific community or the objective scholar who chronicles the accomplishments of a discipline with which he has only fundamental knowledge? When our Country chose the Swedish sociologist Gunnar Mydal to write an objective

report on the status of the American Negro, it clearly valued the view of an outsider. While it is difficult to measure the consequences of either approach, history has shown the value of each. Where the reader feels my own interpretation is only one interpretation that may be made from these historical trends, he or she will no doubt be correct.

II. Where We Are Now: History and Current Status of Evaluation

This section briefly reviews the history of educational evaluation, presents the roles evaluation has traditionally played in education, and summarizes the current status of the field.

Educational Developments and Societal Trends Influencing the Growth and Development of Evaluation.

In the first three decades of this century the measurement of human abilities grew out of early work by Binet, Thorndike, and Thurstone. This newly developed measurement technology had much appeal to educators and was assimulated into educational practice, giving rise to the development of standardized achievement tests which made possible large scale testing programs. The accreditation movement also flourished during this early period and with the development of formal accrediting policies for colleges and schools, program evaluation gained a foothold in education. Later, the Educational Testing Service (ETS) established in 1947 and a national system of research and development centers and laboratories established in 1966 provided additional momentum to the field of evaluation through evaluation projects and contributions to evaluation methodology. (See Borich, 1974, and Poynor, 1974, for a selection of evaluation contributions from these centers and laboratories).

Impact of Operationalism and the Behavioral Objectives Movement*

The concept of behavioral objectives has held a position of importance in the field of evaluation for almost half a century. One origin of the concept of behavioral objectives can be traced to a book by Bridgman (1927) titled the

*I am indebted to Bloom, Hastings and Madaus (1971) for the early origins of this movement.

Logic of Modern Physics. In his book Bridgman pointed to the need to define new constructs by describing the operations used to measure them. Bridgman's concept offered an alternative to the practice of defining constructs by their apparent commonality or lack of commonality with other constructs which, earlier had been defined in the same manner. Through the efforts of Bridgman and parallel efforts of others the idea of operationally defining constructs became incorporated into the behavioral sciences, where constructs such as "motivation," "anxiety," and "learning" were redefined in terms of the measurement operations used to observe them. Other frequently used constructs, such as the construct "insight," took on mostly theoretical significance for lack of practical and reliable means of measuring them. This process of tying construct definition to construct measurement became an integral part of the school known as behaviorism to which the behavioral objectives movements owes its beginning.

The application of operationalism to education resulted in the outgrowth of two distinct but related movements. The first is typified by Tyler's Eight Year Study of Secondary Education for the Progressive Education Association (Smith and Tyler, 1942) in which behavioral objectives were extensively used to evaluate "progressive" attempts to apply new curricula and approaches to instruction. Tyler's contribution is significant not only because it offered the first example of how behavioral objectives could be used to construct evaluation instruments and to appraise the effectiveness of curricula but also because it provided the impetus for many developments in the field which were to follow. Some of the more noteworthy of these were the Taxonomy of Educational Objectives in the Cognitive Domain (Bloom, et al. 1956) and Affective Domain (Krathwohl et al., 1964) and a popular book by Mager (1962) on how to write educational objectives. These volumes, in turn, stimulated an extensive literature on behavioral

objectives, both in support of and critical of their application in the schools (Popham, 1969, Eisner, 1969).

A second movement rooted in a behavioristic philosophy was the programmed instruction and related computer assisted instruction movement of the late 1950's and 1960's. Behaviorally stated objectives were central to both these forms of instruction. The development of programmed and computer assisted instruction depended heavily on the specification and breaking down of content into discrete learnable units having measureable outcomes, for which the concept of behavioral objectives was ideally suited. In this behavioralistic setting, several large development and evaluation projects were begun. Of particular note were evaluations of the Plato and Ticcit computer assisted instruction projects designed to study the cost and effectiveness of computer based instruction for teaching large numbers of geographically dispersed students. (See Alderman, 1978; Murphy and Appel, 1977; and Orlansky and String, 1978, for evaluations of these and other computer based instruction projects.)

The Impact of the Curriculum Reform Movement

A major impetus to the development of evaluation was the curriculum reform movement. Spanning roughly the decades of the 1950's and 1960's the curriculum reform movement was characterized by widespread change in the philosophy, techniques and materials used in teaching elementary and secondary school children. Most notable were the changes which occurred in the sciences shortly after the 1957 launching of the Soviet satelite, Sputnik. Prior to this unsettling event, curricula for the public schools were written primarily by individuals, authoring textbooks which changed only slightly the style and content of earlier versions. Due partly to the inability of any single author to undertake major curriculum reform and partly to the liability to oneself and publisher such reform might present if not saleable, curriculum changes were slow and for the most part conservative. With Soviet competition in the sciences, however, came the impetus for the federal

government to play an increasing role in the field of education, at first through the vehicle of the National Science Foundation and later through the efforts of the U.S. Office of Education and the National Institute of Education. The post-Sputnik era provided the context for new initiatives in the design and development of curricular materials, particularly in the fields of science and mathematics. These initiatives represented not only an effort to reform certain segments of the school curriculum but also to try new approaches to curriculum development which placed decreasing emphasis on the individual author and increasing emphasis on teams of specialists brought together by public monies specifically for the purpose of infusing the school curriculum with the latest scientific advances. New content and innovative ways of presenting it became more palatable with the burden of risk for a development project being shared by teams of specialists sponsored by government monies. Even more appealing was the fact that often extensive discussions, symposia and workshops would accompany these development projects for the purpose of giving teachers and scientists a significant role in the design and selection of content. This unique integration of theory and practice became a key element in a process which was to become characteristic of the curriculum reform movement.

Also of significance was the fact that with the systematic approach to curriculum development the previously isolated concepts of development and evaluation became parts of a unitary process. Because of the experimental nature of much of the content and approaches used, pilot and field testing of instructional components became logical extensions of the curriculum development effort. It was in this context that projects such as the Biological Sciences Curriculum Study (BSCS), The Chemical Education Materials Study (Chem Study), the Physical Science Study Committee (PSSC) and School Mathematics Study Group (SMSG) were

born. These projects contributed significantly to the field of evaluation by employing development strategies which required the repeated testing and revision of components parts of the curriculum. This process of testing well-defined units of a curriculum during development for purposes of revision and modification was later to be coined "formative evaluation" by Scriven (1967). (See Grobman, 1968, for a review of the curriculum reform movement and a history of the Biological Sciences Curriculum Study).

The significant role which evaluation played in these projects stimulated efforts at several universities to mount doctoral training programs in the area of evaluation. Training programs were begun at the Ohio State University influenced principally by Professor Stufflebeam (now at Western Michigan University), the University of Illinois influenced principally by Professor Stake and at the University of Virginia influenced principally by the late Professor Provus. In addition each of these individuals developed in conjunction with his training curriculum an evaluation model which could be used in evaluating educational programs and curricula. These models would later figure centrally in the development of the field of evaluation.

The Impact of ESEA

Despite the influence of the behavioral objectives and curriculum reform movements, there was still relatively little emphasis placed on the evaluation of educational programs by the mid 1960's. It was within this context that the U.S. Congress began debate on the Elementary and Secondary Education Act of 1965 (ESEA). This comprehensive and ambitious educational legislation was to make available large sums of money in the form of grants to universities and local education agencies for educational materials, development and research. As the bill was debated, concern was expressed that there were no assurances

that the federal monies made available would actually result in improvements in the quality of education. This concern was perhaps magnified by the general belief that, in the past, educators had done a poor job of accounting for the federal money they spent.

Motivated by this concern, the Congress insisted on a provision to ESEA requiring that evaluation reports be submitted by grantees reporting the impact of their programs. These guidelines were conveyed to prospective grantees in an ESEA Title III manual published by the U.S. Office of Education, requiring the applicant to:

A. Where applicable, describe the methods, techniques and objectives which will be used to determine the degree to which the objectives of the proposed program are achieved.

B. Describe the instruments to be used to conduct the evaluation, and

C. Provide a separate estimate of costs for evaluation purposes. (p. 48)

Although the final version of the bill did not require evaluation of all the programs (titles) under ESEA, there was a clear mandate from those providing federal funds for education that programs utilizing these funds be accountable for the educational programs, products, and procedures they developed and/or implemented. For the first time educators were required to devote time and resources to evaluating their own efforts.

This emphasis on accountability became evident again in 1971, when a rider was placed on legislation requiring that all ESEA projects be evaluated by the grantee. The current popularity of "sunset" and "sunshine" policies and zero-based budgeting among both state and federal funding agencies reflects this

continued emphasis on accountability. These policies require the recipients of funds to justify refunding of their program each year or program cycle and to make program decisions and expenditures a matter of public record.

Impact of School and Teacher Accountability

The concept of school and teacher accountability emerged as an outgrowth of the ESEA Legislation of 1965 and 1971. Federal agencies and grantees responsible for innovative ESEA programs were only the first to feel the pressure for accountability. Because many of these programs dealt directly with the schools, the accountability demanded of them also raised questions about the school staff who played a prominent role in their implementation. Consequently, teaching effectiveness and the administrative accountability of schools in general often became the focus of attempts to monitor and evaluate federally funded programs. The concepts of "accountability," "cost-benefit," and "quality assurance," filtered down in spirit, if not in substance, to the local school and teacher.

By 1970 community pressures began to bear down on the local school, often demanding accountability in terms of pupil outcome. In some cases school administrators responded to these pressures by concentrating on the more obvious indicators of effectiveness, such as pupil performance on national achievement tests, number of college admissions, and National Merit scholarships. Others began exploring ways to make cost-effective decisions about the operation and management of their school in order to prove that increased revenues actually produced more effective teaching and learning. School administrators embraced accountability procedures in answer to community pressures for more objectively determined and effective ways to spend school revenue and to make internal decisions that could be defended to school boards, PTA's and professional groups.

It was within this context of widespread community concern about higher but

apparently unproductive school expenditures that some state governments began discussing legislation requiring the appraisal of school-district personnel. A prime example of state-enacted accountability legislation was California's Stull Act passed in 1971, requiring that school boards in that state evaluate their educators yearly and provide recommendations for their professional development. The Stull Act gave local communities a mandate to develop procedures for appraising school district personnel and for periodically reporting appraisal data back to the teacher in order to upgrade his or her performance. A major impact of the school and teacher accountability movement on the general field of evaluation has been in the area of process evaluation. In order to evaluate the performance of teachers, reasearchers have operationally defined a large number of teacher behaviors or "competencies" which have shown to relate to pupil achievement. Many of these teacher behaviors and related instrumentation have been used by evaluators to study the processes with which instructional staff implement educational programs and curricula. (See Borich, 1977, for other contributions of the school and teacher accountability movement.)

A summary of the contributions to evaluation associated with operationalism, curriculum reform, ESEA Legislation and school accountability appears in Table 1.

Table 1

Some Contributions Associated with

Four Milestones in the Field of Evaluation

Milestones	Contributions
1. Operationalism	Defininiq constructs by the procedures used to measure them
	Use of behavioral objectives for program design and evaluation
	Programmed instruction

Table 1
(continued)

(1. Operationalism)	Computer assisted instruction
2. Curriculum reform	Increased federal expenditure in education
	New initiatives in instructional techniques and materials
	Cooperation of scientists and teachers on the design of curricula
	Integration of curriculum development and evaluation as a unitary process (formative evaluation)
	Doctoral training programs in evaluation
3. Elementary and Secondary Education Act of 1965	Federal commitment to evaluation
	Federally mandated and funded evaluations
	The principle of refunding contingent on evaluation results
	Project accountability at the local level
4. School accountability	Teacher and administrator accountability
	Pupil behavior as criterion of program (teacher) success
	State mandated evaluations
	Process evaluation techniques and instruments
	Evaluation as feedback for professional development.

Response to the Demand for Effective Evaluation

Although citizens were generally positive about the explicit mandates contained in ESEA legislation and California's Stull Act, it became evident by mid 1970 that educators were not prepared to effectively implement either of these new mandates. Moreover, the sudden increase in demand for capable evaluators brought about by these mandates quickly exhausted the supply. Few educators had any formal training in evaluation and often local school personnel were pressed into service as program evaluators.

One obstacle to the implementation of these mandates was the inability of local, state and federal administrators to apply the mandates. The evaluation concepts created by educators in the preceding decade no longer seemed adequate to answer the questions which now were being asked of these programs. After reviewing the evaluation reports of ESEA programs, Guba (1969) concluded that,

> The traditional methods of evaluation have failed educators in their attempts to assess the impact of innovations in operating systems. Indeed, for decades the evidence produced by the application of conventional evaluation procedures has contradicted the experiential evidence of the practitioner. Innovations have persisted in education not because of the supporting evidence of evaluation but despite it. (p. 28)

and at another point argued that,

> When the evidence produced by any scientific concept or technique continually fails to affirm experiential observation and theory arising from that observation, the technique may itself appropriately be called into question. (p. 30)

With the emergence of ESEA came not only a need for new management strategies to monitor these programs but also a need for improved evaluation designs to test their effectiveness.

Reflecting on the current state of evaluation practice the report of the Phi Delta Kappa (PDK) national study committee on evaluation (Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman and Provus, 1971) concluded that evaluation was "seized with a great illness" (p. 4). The "symptoms" of this illness, as stated by the PDK committee were:

(1) The Avoidance Symptom - Evaluation is perceived as a painful process which may expose a school districts' programs or individuals' shortcomings. Evaluation is avoided unless absolutely necessary.

(2) The Anxiety Symptom - Evaluation evokes anxiety. The educator as well as the evaluator knows how cursory, inadequate, and subject to error the evaluation process can be. The ambiguity in the evaluation process engenders anxiety in both the educator and evaluator.

(3) The Immobilization Symptom - Despite federal requirements to evaluate, evaluative data on educational programs, products and procedures are still rare. This lethargy and lack of responsiveness is symptomatic of deeper ills.

(4) Lack of Theory and Guidelines Symptom - There is a lack of unified theory of evaluation. With evaluators differing among themselves about what evaluation should and should not be, the evaluator in the field is left to his own

devices for conducting evaluative inquiry; there are few useful guidelines for him to follow.

(5) The Misadvice Symptom - There is ample evidence that evaluation consultants have provided educational practitioners with poor advice. Not only is there a lack of adequate guidelines but obtaining advice from an evaluation "expert" is no guarantee that a technically sound evaluation report will result.

And, to these were added the lack of trained personnel, the lack of knowledge about decision processes, the lack of values and criteria for judging evaluation results, the need to have different evaluation approaches for different types of audiences, and the lack of techniques and mechanisms for organizing, procuring and reporting evaluative information.

The foregoing suggest that at the beginning of the past decade the relatively new discipline of evaluation was indeed besieged with problems which could be conceptualized as deficiencies. These deficiencies, though, were themselves symptoms of a more fundamental ill: the lack of an adequate definition of evaluation and the lack of adequate evaluation theory.

Traditional Definitions of Evaluation

The lack of an adequate theoretical base for the discipline of evaluation has often been cited as a factor which has stifled the development of the field and its ability to provide meaningful evaluative data to educational practitioners. Even more problematic, however, is the lack of consensus among evaluators as to how evaluation should be defined.

Evaluation has been arbitrarily defined in a number of ways. Four definitions which have achieved some popularity during the development of the field are the following.

Evaluation as measurement. This early definition of evaluation came to the forefront during the 1920's and 1930's with the rise of the measurement movement in psychology and education. Evaluation received considerable impetus from the emergence of the science of measurement and it is not surprising that the terms were equated during the 1930's. More current measurement definitions have been expanded to give a broader focus to the term evaluation but maintaining the close tie to measurement. Consider the following definition from a measurement text by Thorndike and Hagen (1965, p. 27):

> The term "evaluation" as we use it is closely related to measurement. It is in some respects more inclusive, including informal and intuitive judgments....saying what is desirable and good. Good measurement techniques provide the solid foundation of sound evaluation.

Defining evaluation as measurement has the advantages of building directly on the scientific measurement movement with its attendant objectivity and reliability. Further, measurement instruments yield data which are mathematically and statistically manipulatable, facilitating the establishment of norms and standards. The disadvantage of this definition of evaluation is that it is totally dependent on the development, administration, scoring and interpretation of measurement instruments (tests, questionnaires, attitude scales, etc.) which take time to develop and are relatively expensive. This approach also obscures judgments and judgment criteria. Scores become entities

unto themselves while concepts behind the scores tend to be obfuscated. A final disadvantage, and perhaps the most important, is that variables which do not lend themselves readily to measurement are often eliminated or ignored. (See Thorndike and Hagen, 1965, and Ebel, 1965, for further explication of this approach to evaluation.)

Evaluation as determining congruence. This widely accepted definition of evaluation is concerned with the congruence between performance and objectives, i.e., determining the degree to which the performances of students are congruent with the objectives of instruction. The major proponent of this definition was Tyler who, reporting on his Eight Year Study of Progressive Education (Smith and Tyler, 1942), viewed educational objectives as changes in behavior. If a program succeeded in bringing about the desired changes (i.e., if there was a congruence between student performance and the objectives) then the program was judged successful.

A major advantage of this approach is that it forces educators to conceptualize clearly the goals of instruction and requires their full articulation. Further, this emphasis on objectives provides at least implicit criteria for judging the success of a program. Another distinct advantage of this definition is that it allows for the evaluation of education processes (e.g. teacher behavior) as well as educational products (e.g. student achievement).

One disadvantage of this definition includes the fact that objectives have to be made specific to be measurable, which may obscure important but less specifiable objectives intended by program developers. Another disadvantage is the heavy emphasis placed on student behaviors. A new staffing policy or instructional strategy is evaluated in terms of student achievement, and such issues as cost-effectiveness,

teacher satisfaction and student discipline may be ignored. A related disadvantage of emphasizing student achievement is that congruence evaluations tend to be ex post facto. Although Tyler's approach allows for evaluation of process, the data emphasized in this approach, that of student performance, are available only at the end of the project when the performance of students is compared to program objectives. Thus, valuable process data are often not collected (or at least not emphasized) and the opportunity for feedback and program modification is often lost. (See Tyler, 1950, and Furst, 1964, for a further discussion of this definition of evaluation.)

Evaluation as professional judgment. The definitions discussed above place little emphasis on the judgmental process. Attaching value to the data was assumed. In this definition evaluation is professional judgment. The most common practice in this approach is site visitation, such as that used in accrediting schools and colleges. A visiting team of experts come to "soak up" the environment, and to use their expertise in rendering a judgment of program effectiveness.

Advantages of this approach include ease of implementation, consideration of a large number of quantitative and qualitative variables (including the context, experience and expertise of the evaluators) and quick "turn around" of results and conclusions. Major disadvantages include the questionable objectivity and reliability of the judgments that are made, the ambiguity of the judgment criteria, and the difficulty in generalizing results of the evaluation to other programs or institutions.

Evaluation as applied research. Although evaluation usually has not been defined in terms of research, a sorting through of evaluation studies reveals a strong reliance on the scientific method and an even heavier emphasis on

the experimental designs and statistical tools of research. This result is not surprising when considering that the typical evaluator is usually extensively trained in the methodology of research and often only minimally trained in those concepts unique to evaluation.

Despite obvious advantages of classical research methodology, such as experimental control over variables and the statistical power of parametric statistical techniques, there are practical considerations which limit the applicability of these procedures to educational problems. These were presented by Stufflebeam et al., (1971) and are updated and summarized below with some extensions and modifications.

1. Laboratory antisepsis. Cooley and Lohnes (1976) point out that scientific research attempts to validate the existence of cause-and-effect relationships with the ultimate goal being the development of a consistent and parsimonious theory of natural phenomena. Evaluation research, on the other hand, is concerned with means-end relationships with the ultimate goal being a rational choice between alternatives for action. Because scientific research pursues universal laws, knowledge must be obtained in a context-independent way. Experimental manipulation is used to control all confounding and extraneous variables. The evaluation of an educational program is concerned, however, with all the mitigating variables affecting some educational outcome. "In order to provide useful data, educational evaluation does not need the antiseptic world of the laboratory, but the septic world of the classroom

and school" (Stufflebeam, et al., 1971, p. 22). Laboratory research designs require conditions usually not attainable in evaluation contexts.

2. Effects of intervention. In scientific research, variables are manipulated by the experimenter to create critical comparisons of the ways variables interact. Thus, the experimenter's intents become part of the data. The evaluator, on the other hand, attempts to assess interactions in a real rather than contrived environment. His data collection must be done unobtrusively so as to not confound his results.

3. Terminal availability of data. Research designs typically attempt to assess the effect of some experimental treatment. The treatment is administered, then data are collected and analyzed. Data for making judgments are available only after the treatment has been administered. This precludes the use of data to refine a treatment, although continuous refinement of an ongoing educational program is a frequent function of evaluation.

4. Single treatments only. For purposes of experimental control, scientific research requires that a treatment be evaluated alone. If several treatments are operating simultaneously, their effects will confound each other. Educators, on the other hand, cannot withhold a potentially beneficial educational program because students are concurrently enrolled in other treatments.

5. Effects of control variables. Random assignment is generally not possible in educational settings. Thus, to equate treatment groups (in order to enhance their comparability) evaluators usually match groups on selected control variables such as intelligence levels, ethnic mix, classroom size, socioeconomic status, and the like. The problem with this procedure is that criterion variables (such as measures of cognitive or affective achievement) are often correlated with these control variables causing treatment differences to be obscured.
6. Inapplicability of assumptions. Some assumptions underlying the use of parametric statistical procedures may not be met in the usual evaluation setting, for example when distributions are severely skewed, relationships nonlinear, or group variances unequal.
7. Restricted decision rules. Conventional statistical techniques contain decision rules of the simple "go-no go" variety. A null hypothesis may be rejected or accepted or treatment X may be judged better than treatment Y. Evaluators are often asked to bring their expertise to bear in more complex decision settings.

In a similar fashion Hemphill (1969) has distinguished research from evaluation along six dimensions: problem selection, replication, determination of data, determination of hypotheses, values and control. To emphasize the differences between research and evaluation, Hemphill cast these dimensions in parallel form. These dimensions are noted in Table 2.

Table 2

Contrasts Between Research and Evaluation

	Research	Evaluation
1.	Problem selection and definition is the responsibility of the individual doing the study.	Many people may be involved in the definition of the problem and because of its complexity, it is difficult to define.
2.	Given the statement of the problem and the hypothesis, the study can be replicated.	The study is unique to a situation and seldom can be replicated, even approximately.
3.	The data to be collected are determined largely by the problem and hypothesis.	The data to be collected are heavily influenced if not determinted by feasibility.
4.	Tentative answers may be derived by deduction from theories or by induction from an organized body of knowledge.	Precise hypotheses usually cannot be generated; rather the task becomes one of testing generalizations some of which may be basically contradictory.

5.	Value judgments are limited to those implicit in the selection of the problem.	Value judgments are made explicit by the selection and definition of the problem as well as by the development and implementation of the study.
6.	Relevant variables can be manipulated or controlled by including them in the design.	Only superficial control of potentially confounding variables can be achieved.

While all-exclusive distinctions between research and evaluation are often subjects of controversy, most evaluators and researchers implicitly support a broad separation between these two modes of inquiry. So sharply has the line between research and evaluation been drawn at times that some evaluators contend that the two modes of inquiry are basically incompatible and ultimately must employ different methodology.

Models for Evaluation

Different conceptions of evaluation have spawned numerous paradigms or models for implementing an evaluation study. These paradigms or models, however, represent different conceptions of evaluation more than they do different objectives or contexts for evaluation. Matching a particular type of evaluation problem to a particular model does not seem possible nor does there seem to be explicit rationale as to why an evaluator might choose one model over another. This has left evaluators without criteria for selecting the most appropriate model for a given evaluation problem.

Some educators and program develops operate under the assumption that a

variety of specific evaluation models exist which are readily applicable to their particular educational problem. When the time for evaluation comes, the task is deemed as a simple one of selecting an appropriate model, plugging the program into it and analyzing the results. Evaluation models generally are not precise or specific and the choice of an evaluation model is itself a value judgment about how an educational program should be evaluated. An alternative to selecting a general model is to adapt a model developed for a specific setting and generalize it to one's own problem context. Highly specific models are, however, developed within a narrow context and are generalizable only to those settings which have identical or highly similar administrative organizations, funding and political presses, personnel compositions, data analysis support systems, client populations, educational objectives, and personnel biases about what is and is not important in evaluating a program.

A variety of evaluation models abound in the professional literature. Some are purposively general so as to be applicable to a variety of educational contexts (cf Hammond, 1973; Metfessel and Michael, 1967; Provus, 1971; Stake, 1967; Stufflebeam et al., 1971), while others are developed to meet evaluation needs in a specific setting (cf Belliott, 1969; Dykstra, 1968; Emrick, Sorenson, Stearns, 1973; and the Interservice Procedures for Instructional Systems Development (IPISD) model, 1975). To underscore their general nature three popular evaluation models are summarized below.

The Discrepancy Evaluation Model

This model, developed by Provus (1971), divides evaluation into five stages.

Stage 1. This stage documents program description. The evaluator obtains from the program staff a comprehensive description of program inputs, processes and outputs. These are compared to the staff's definition of the program. Discrepancies are noted and used to modify program definition such that it is congruent with program components.

Stage 2. In this stage field observations are used to determine if the program is being implemented as intended. Discrepancy information is used to modify program implementation. This is also called "process" evaluation.

Stage 3. In this stage it is determined whether program components are engendering the attainment of intermediate or enabling educational objectives as intended. It is a check on whether student behavior is changing as expected. Discrepancy information is used to modify either the program components or the objectives. This stage is similar to Scriven's (1967) concept of formative evaluation.

Stage 4. In this stage it is determined whether program components are leading students to terminal program objectives. This stage often uses pre-post behavior change and sometimes control vs. experimental comparisons. This stage is similar to what is called summative evaluation.

Stage 5. In this stage (which is not always applicable) the experimental program is compared to a realistic alternative. An experimental or quasi-experimental design (Campbell & Stanley, 1966) is used to prove that program benefit is commensurate with cost.

This model's components include agreeing on program standards, determining whether a discrepancy exists between aspects of a program and standards governing those aspects, and using discrepancy information to identify program weaknesses. Discrepancy information at each stage leads to a decision whether to proceed to the next stage or to alter either program standards or operations. Advancement to a subsequent stage is contingent on attaining congruence between operations and standards at the previous stage. If congruence is not possible program termination is recommended, although in practice this option is rarely chosen.

The Stake Model

This model (Stake, 1967) divides educational programs into three major concepts:

1. Antecedents - conditions existing prior to training that may be related to outcomes such as previous experience, interest and aptitude.
2. Transactions - encounters of students with teacher, author with reader, parent with counselor or some educational activity such as the presentation of a film, a class discussion or working a homework problem.
3. Outcomes - measures of the impact of instruction on students, teacher, administrators, parents or others. These are usually measures of abilities, achievements, attitudes, aspirations, etc. Outcomes can be immediate or long range, cognitive or affective, personal or community-wide.

To Stake, evaluation involves (1) examining the logical contingencies that exist between intended antecedents, transactions and outcomes; (2) determining the congruence between intended and observed antecedents, transactions and outcomes; and (3) determining the empirical contingencies between observed antecedents, transactions and outcomes. Illogical contingencies, lack of congruence, and possibly, a failure to establish empirical contingencies aid in identifying program weaknesses.

The CIPP Evaluation Model

This model, developed by the Phi Delta Kappa Commission on Evaluation (Stufflebeam, et al., 1971), divides evaluation into four distinct strategies

- Context evaluation, Input evaluation, Process evaluation and Product evaluation, thus the acronym CIPP. Context evaluation has as its objective to specify the operational context and to identify problems underlying needs. Input evaluation is concerned with identifying and assessing system capabilities. The objective of process evaluation is to identify defects in procedural design or implementation and to document project activities. The goal of product evaluation is to relate outcome information to objectives and to context, input and process information. If these relations are not specifiable, program weaknesses are suspected.

As can be noted from this overview, evaluation models represent very general aids or hueristics to conceptualizing evaluation designs. Other more technical models embodying more specificity have been developed for highly specialized, idiosyncratic applications but these have limited generalizability across educational settings.

Models as Heuristics

The desire among evaluators to identify models is understandable, since there is the hope that once these models are established they can be used in a large variety of evaluation contexts. However, evaluation does not work that way. The techniques and methods brought to bear in an evaluation are a function of the problem, the clients for whom the evaluation is being conducted and the amount of time and money which can be devoted to it. While evaluation is certainly not an art form, evaluation models can communicate only a relatively small set of categories and constructs that might be useful in planning an evaluation. Thus, in part, the problem of choosing the correct evaluation model derives from a somewhat natural tendency to see an evaluation

model as more than it is - as a methodology for actually conducting the evaluation instead of a meta-methodology or framework into which must be plugged more specific constructs and methods.

This false expectation for evaluation models has been known to lull educators into not giving much thought to the evaluation process. Further, when it is discovered that there is not a "tight" evaluation model available to provide needed evaluative data, it is often considered a shortcoming of the evaluator. Good evaluation procedures require the input of evaluation specialists early on in program planning and development. The frustrations of educators over evaluation often stem from their own failure to consider evaluation issues throughout the educational program development process. Surprising to some is the fact that an evaluator is not an all-knowing guru with a magical bag of tricks (models) that will compensate for the failure to properly consider and plan evaluation activities early on in program planning and development. Evaluation models do not provide answers but do provide useful guidelines or heuristics which can help organize thinking about how an evaluation should be conducted. This heuristic role for models, which has not always been appreciated in evaluation theory or practice, has been described by Kac (1969):

> The main role of models is not so much to explain and to predict--though ultimately these are the main functions of science--as to polarize thinking and to pose sharp questions. Above all, they are fun to invent and to play with, and they have a peculiar life of their own. The "survival of the fittest" applies to models even more than it does to living creatures. They should not, however, be allowed to multiply indiscriminately without real necessity or real purpose.

Commonalities Among Models

A common approach to obtaining useful evaluation data has been to develop one's own model borrowing, where appropriate, from existing models in the literature, thus avoiding the model selection problem. This is generally the preferred alternative for educators who wish to assure the best match of program purpose and context to an evaluation model. Table 3 indicates the fundamental evaluation concepts shared by these three models and where they tie into each model. Table 4 provides a specification matrix indicating how each model addresses major considerations in chosing an evaluation design. Taken together these tables provide a means of judging the applicability of the component parts of these models to specific evaluation contexts. Further, they serve to illustrate the commonalities and distinctions encountered in studying established, well-known evaluation models. However, evaluators must be mindful that some solutions which are suitable for models may not apply to the real world. Evaluators must never forsake the real world for the complexity of models which purport to describe it.

TABLE 3

Some Commonalities Among Three Evaluation Models*

General Concept	Provus	Stufflebeam	Stake
Transaction	Transaction		Transaction
Enabling Behavior	Enabling	Instrumental	Immediate
Input Evaluation	Stage 1	Stage 2	Antecedents
Product Evaluation	Stage 4	Stage 4	Outcomes
Process Evaluation	Installation Stage	Process Stage	Congruency
Program Definition	Program Definition Stage	Input Stage	Logical Contingency
Standards	Each Stage		Relative, Absolute
Objectives	Program Definition		Intents
Judgment	Stages 1-5		After Description
Context		Context	
Antecedents			Antecedents

*Wording used by these authors appears in their respective columns.

TABLE 4

Characteristics of Three Evaluation Models

Characteristic *	CIPP	Stake	Provus
1. Purpose of evaluation	to make better, more defensible decisions	to describe and judge the merit of a thing	to uncover discrepancies between standards and performance
2. Implied role of evaluator	information provider, serves the decision-maker	makes judgments about the effectiveness of a program from descriptions & standards	compares standards with performance at various stages to revise or terminate program
3. Relationship to objectives	high	high, "intents" are objectives	high "standards" are objectives
4. Types of evaluation activities proposed	context, input process, product	description, judgment, logical & empirical contingency, congruency	program definition, installation, process, product, cost/benefit
5. Unique constructs	context	logical contingency	discrepancy

*Constructs in this column were selected from Worthen and Sanders (1973).

TABLE 4
(continued)

	Characteristic	CIPP	Stake	Provus
6.	Relationship to decision maker	integral	unclear	high
7.	Some criteria for judging evaluations	Did the evaluator collect context input, process & product data?	Did the evaluator look for logical contingencies & collect judgement data?	Did the evaluator collect data and check for discrepancies within each stage?
8.	Implications for evaluation designs	mostly qualitative decisions except for product evaluation, where a control group is applicable	deals mostly with descriptions and judgments. Control group helpful but not necessary, judgments can be absolute	comparisons between standards & performance at each stage are essential, control group is needed for cost/benefit stage.

III. Prospects for the Immediate future: Emerging trends in educational evaluation.

Since the initial ESEA legislation of 1965, some evaluators (Apple, 1974; Cooley & Lohnes, 1976; Kaufman, 1972; Guba, 1978; Provus, 1971; Scriven, 1973; Stake, 1967, 1970; Stufflebeam, et. al., 1971) have attempted to provide a stronger basis for evaluative theory and in so doing have implicitly or explicitly offered new definitions and theoretical bases for evaluation. These new conceptualizations build on previous ones and can be broken into four types or styles of evaluation: decision-oriented evaluation, value-oriented evaluation, naturalistic evaluation and systems-oriented evaluation:

Decision-oriented Evaluation

The PDK National Study Committee on Evaluation (Stufflebeam, et al. (1971) defines educational evaluation as:

> ...the process of delineating, obtaining, and providing useful information for judging decision alternatives (p. 40).

Provus (1971) similarly defines evaluation as:

> ...primarily a comparison of program performance with expected or designed programs, and secondly, among other things, a comparison of client performance with expected client outcomes (p. 12).

It can be seen that these decision-oriented definitions are heavily influenced by Tyler's (1950) congruence definition of evaluation but are of a much broader scope and are oriented toward a decision tree logic. Inherent in this approach is an emphasis on comparing "what is" with "what should be" and using discrepancy data as a basis for decisions.

The major advantage of this approach is that by following the models associated with these definitions an evaluator is better able to provide the kinds of information desired by decision-makers. Acceptance of the decision-oriented stance requires that clearly defined goals and objectives be elucidated prior to the collection of data, thus ensuring the presence of adequate criteria for judging the adequacy or relative merit of a program. The presence of prespecified criteria for judging program effectiveness may, however, be a disadvantage as the following discussion notes.

Value-oriented Evaluation

Some authors in the field of evaluation have taken exception to the notion of decision-oriented evaluation and its implications for the conduct of evaluative research. The primary criticism of decision-oriented definitions is that evaluation is viewed as a shared function. The role of the evaluator is to provide a decision maker with meaningful information; the decision maker makes the actual judgment of value or merit.

A value-oriented definition of evaluation stresses the value-judgments made in evaluating educational programs and describes the act of judging merit or worth as central to the role of the evaluator. Worthen and Sanders (1973) define evaluation as "...the determination of the worth of a thing." (p. 19). Scriven (1967) considers the evaluator who does not participate in the decision making process as having abrogated his role. Stufflebeam et al., (1971) and Stake (1967) argue that by participating in decision making, the evaluator loses his objectivity and hence, his utility. Differences in these approaches are more than semantic for they imply different evaluation activities.

Within the decision-oriented approach, the evaluator is dependent upon the decision maker for the way the decision context is to be defined and for the values and criteria that are to be used to judge program success (these are usually termed program intents, goals, or purposes). Cooley and Lohnes (1976) and Apple (1974) point out that there is no evidence to suggest that the decision maker is any more capable than the evaluator to define decision settings, alternatives, and values. Indeed there may be (and often are) social, institutional, and political presses on the decision maker which may lead him to opt for evaluation procedures that skirt or ignore key evaluation issues. Apple (1974) makes the case that decision-oriented evaluation is a conservative practice not conducive to the acceptance of educational innovation but rather supportive of the status quo. Apple's point is that the limits of the decision-oriented evaluators work is circumscribed largely by the already developed program, and therefore, the evaluator cannot deal with the issues, concerns and objectives which predate the program and to which the program is supposed to be responding. Once the program is in place, the evaluator's role is to work with it (i.e. revise or modify it) regardless of whether it is the best means to the desired end.

Scriven (1974) argues that value judgments are a crucial part of all sciences, particularly methodological value judgments, and there is no reason to dismiss them in evaluation. He calls for goal-free evaluation, insisting that all aspects of an educational program should come under the scrutiny of the evaluator and that nothing should be taken as given from the client or agency soliciting evaluation expertise. The following illustrates his point:

> The goal-free evaluator is a hunter out alone and goes over the ground very carefully, looking for signs of any kind of game, setting speculative snares when in doubt. The goal-based evaluator, given a map that, supposedly, shows the main game trails, finds it hard to work quite so hard in the rest of the jungle.
>
> (Scriven, 1973, p. 327)

Scriven argues that while knowledge of goals is necessary for effective planning and implementation it is unnecessary in evaluation and may even blind the evaluator to important program effects.

Scriven (1973, 1974) and Apple (1974) also emphasized the social responsibility of the evaluator. Scriven offers the hypothetical example of an educational program aimed at increasing self-sufficiency. After some evaluative activity the evaluator discovers that in addition to fostering self-sufficiency, the program engenders contempt for the weak, sick, old and congentially deformed. Scriven contends that these findings should count against the program although the program developer might be concerned only with the achievement of his announced and intended goal. The welfare of the consumer (usually in the case of education, society as a whole) is considered a proper concern of the evaluator.

Apple puts forth a similar argument:

> The tendency in the face of the all-too-usual finding of "no significant difference" is to argue for better teacher training, for better instructional materials, for more sophisticated administrative systems designs and the like. However, it may well be that more basic questions must be

> asked, that even the obligatory nature of the institution of schooling may need questioning, or that educators are asking the wrong kinds of questions...much low achievement on the part of the students could be attributable to a symbolic dismissal of school itself as a meaningful institution...unresponsive to human sentiments...Educational problems are considerably more fundamental than educators may suppose, and it places responsibility on the individual educator to examine his or her own professional activity in a wider social and political context. (Apple, 1974, pp. 28-29)

The implication of Apple's view for the evaluator has been elucidated by Becker (1974), a sociologist, who forshadows how the evaluator who fails to give deference to the status quo is likely to be received by the decision maker:

> For a great variety of reasons, well-known to sociologists, institutions are refractory. They do not perform as society would like them to. Hospitals do not cure people; prisons do not rehabilitate prisoners; schools do not educate students. Since they are supposed to, officials develop ways of denying the failure of the institution to perform as it should and explaining those failures which cannot be hidden. An account of an institution's operation from the point of view of subordinates therefore casts doubt on the official line and may possibly expose it as a lie (Becker, p. 113).

Becker believes that any approach the sociologist or evaluator might take is inherently value ladden and will implicitly support either the subordinate (program participants') or superordinate (program manager's) point of view. While this may be true, Becker's comment also raises the possibility that due to the efforts of decision makers to protect the status quo or allow only changes to be made that are congruent with the existing social, political and organiza-

tional structure, evaluators may be implicitly designing evaluations that examine only the efficacy of the program manager's or participant's point of view, avoiding all other points of view. Such a design is most likely when the goals and objectives for a program must be taken as "givens" and the evaluation designed around them.

Dewey's Conceptualization of Valuation

A cohesive value-oriented theoretical perspective on evaluation has recently been put forth by Cooley and Lohnes (1976). Their stance is based on the early work of John Dewey (Dewey, 1922, 1939) and borrows from Handy's work on the study of values in the behavioral sciences (Handy, 1969, 1970; Handy & Kurtz, 1964). While the propositions of Cooley and Lohnes' theory of valuation are quite similar to and generally subsume those of Apple, Scriven, Worthen and Sanders, and others, they are put forth in a more direct fashion that have practical implications for some additions to evaluation methodology.

They assert that the value statements inherent in educational programs can themselves "...be analyzed into a set of propositions subjectable to empirical investigation and that failure to perform such analyses in evaluation studies is inexcusable" (Cooley and Lohnes, 1976, pp. 9-10). They argue that the values which have guided educational practice have traditionally been determined by politics and custom and that their validity has not been challenged by educational researchers. They find it curious that value propositions have evaded empirical scrutiny despite educational researchers' heavy emphasis on empiricism. Clear thinking about values in education is considered essential because educational practice is generally influenced by the value attached to desired educational goals. The alternative to rational inquiry into values is the determination of values on the basis of power which places the educational enterprise "...at the mercy of special interest groups who commend values favorable to themselves as universals" (p. 10).

A basic premise of Dewey's notions about values and valuations was that values could be mistakenly viewed as absolutes only if they were considered out of context. When considered in context, values lend themselves to elucidation as propositions about real entities (matter of fact) and the error of ascribing to them absolute or universal properties is thus avoided. The task of the evaluator becomes one of ascertaining whether value propositions inherent in an educational setting reflect only convention or tradition or whether they imply empirically testable relationships between educational means and ends.

Consider the hypothetical example in which an evaluator is called in to determine whether an inservice training program for teachers would increase the teachers' appreciation of the difficulties encountered by Spanish-speaking children in a predominantly English-speaking community. The foregoing discussion suggests that the evaluator should consider the context before proceding. Did school administrators merely assume that a general inservice program would have this effect? Was pressure applied to administrators to improve teacher understanding of cultural differences? Was the program developed just because funds were available? Or because it was politically expedient for an elected school official? Or was it because a survey of teachers, parents, and students indicated that such an inservice program would be beneficial? The latter possibility is desirous but seldom encountered.

The value judgment explicit in the above example is that teachers need to have a better appreciation of the educational difficulties encountered by Spanish-speaking children. Also implicit is that the teachers presently are insensitive to these problems, that these students are being shortchanged in

their education and that the administration is quite concerned over this state of affairs. Each of these value propositions may or may not be true and is capable of being empirically determined.

Optimally, the need for such a program would be ascertained before it is developed and implemented. However, this is not always done. Evaluators are usually ignored in program planning, development, and often in implementation. This greatly limits the evaluation expertise that could be brought to bear in the educational setting. Evaluation has much to offer in terms of the "front end" work of educational programming and significant inroads have been made in the area of needs assessment (see Kaufman, 1972, 1976, 1977). This issue will be discussed subsequently, but let it suffice to say here that the notions of Dewey (1939), particularly as they are elucidated by Cooley and Lohnes (1976), provide theoretical justification for the involvement of evaluators early on in an educational endeavor generally, and for the conduct of needs assessments particularly.

Another significant aspect of Dewey's theory of valuation is that he made no absolute distinction between means and ends. Any educational event or condition (e.g., a particular teaching strategy, student achievement in a particular area, etc.) can be viewed as occupying space on a continuum such that it is simultaneously an end to those events and conditions that preceded it and a means to those that follow. For example, inservice education is a means to improved teacher performance which in turn is a means to successful educational settings, etc. Dewey (1922) makes the further assertion that it is only when an end is conceptualized as a means is it fully understood, appreciated or even

obtainable.

To some extent evaluators have taken means-end relationships into account by dividing outcomes into enabling, those that are prerequisite to the attainment of terminal outcomes and terminal outcomes, those that are expected at program completion. Provus (1971) carried the means-end continuum one step further by articulating the concept of ultimate outcomes, those that are expected sometime after program completion. For the evaluator following Provus' model, terminal outcomes are also enabling in that they, too, become means to still other, ultimate ends. Cooley and Lohnes (1976) have argued that there can be no ultimate outcomes unless one appeals to some higher order good. Hoban (1977) suggests that these higher order ends might be chosen from among the values shared by our society such as affection, enlightenment, rectitude, respect, skill, power, wealth and well-being, concepts with which a philosopher not an evaluator would be comfortable. Yet, it would be admirable for the evaluator to make explicit the means-end relationship which is implicit in every evaluation setting, testing its logic and direction against some acknowledged higher-order good at least one step up on the means-end continuum.

The immediate problem for the evaluator is one of determining where to break into the means-end chain for purposes of data collection. Infinite regress is possible in either direction. Cooley and Lohnes suggest that focusing on the present resolves the dilemma. By striving to endow "...present educational policies with a more unified meaning" (p. 13), the evaluator establishes for himself a bounded context for his evaluative activities. This context cannot, however, be too restrictive. Judging the relative value of several competing ends, for example, is very much influenced by each alternative's role as a means to subsequent ends and it is through this logic that relative judgments of worth can be made.

> The question is always what kind of world we want. It is never the narrow one of how to maximize some fixed type of gain...The very important principle is that clarification and transformation of aims or goals of education will be a result of, not a prerequisite for, evaluation research. (Cooley and Lohnes, p. 14)

This approach to evaluation also emphasizes that both means and ends are subject to judgments of value. This position is similar to Scriven's (1967) concepts of formative and summative evaluation. Because the differences between means and ends are seen as superficial, this theory of evaluation poses no restrictions on the evaluation activities that may be pursued in either the formative or summative mode and argues that both modes be utilized.

Another relevant point made by Cooley and Lohnes is that evaluation should not be conceptualized as a single product (usually a monograph) delivered at the conclusion of an evaluation. Rather, it should be viewed as a process in which the evaluator interacts with all other interested parties for an extended period of time. This allows for resolution of differences in opinion, viewpoint, and interests. Cooley and Lohnes consider this version of educational evaluation as "...a process of conflict resolution through intelligent social deliberation" (p. 16). This approach suggests an interactive mode which allows the emergence of a common conceptualization of the educational program among all involved parties and fosters a consensus of program need, design, implementation, and evaluation.

This version of evaluation also stresses the education of all persons involved in evaluation. Put simply, the evaluation should be a learning experi-

ence for all involved. It is not unrealistic to expect that the various parties to an educational endeavor should come to understand more precisely what they are trying to do and why, how their educational programs achieve the results they do, and how each participant may individually facilitate the attainment of successively higher level educational ends in a meaningful way. Thus, evaluation may be viewed as an educational procedure (or means) itself which has as its potential ends-in-view a more harmonious, pleasant and effective educational setting.

A final point about the value-oriented approach to evaluation is that it is inherently humanistic. Educators who consider themselves in the humanistic camp would be attracted to the value-oriented approach because it focuses on the total effects of a program and short and long range outcomes as part of a larger means-end continuum. Also, emphasis is placed on the empirical validation of goals and values, thus preventing them from being determined arbitrarily. The conceptualization of a means-ends continuum provides a foresightful vision of ultimate program effects. The goal free bias inherent in the approach provides a rationale for being sensitive to unknown or unintended program effects. This theoretical view of evaluation has the potential for breaking the traditional mental set of evaluation and provides evaluators with a framework for providing information which can be used to reduce undesirable conditions in society, such as illiteracy, anomie, crime, and racial strife, when the amelioration of these conditions is stated as a higher order end. Cooley and Lohnes state:

> What has been missing in controversies over the schools is convincing evidence which relates choices of educational practices to ends which society values, ends which satisfy needs. Generating such evidence is what evaluation is all about. (p. 18)

Cooley and Lohnes (1976) rediscovery and updating of Dewey's principles of valuation represents a significant addition to evaluation. Primarily, it provides logical and theoretical justification for evaluation concepts, designs, and activities recently called for by other authors in the field (Scriven, 1967, 1973, 1974; Apple, 1974; Worthen and Sanders, 1973; Kaufman, 1972, 1977; Borich, 1977). This justification has been sorely lacking. Unguided by a prudential theoretical basis, evaluation has moved in directions not always conducive to the ultimate improvement of educational quality. Secondly, this theoretical perspective has practical implications for the ways evaluation should be conducted which are in some ways at variance with traditional approaches. Inherent in this perspective is a call for new methods and new concepts in the field of evaluation leading to a considerably expanded and more flexible role for the evaluator. Lastly, this approach is ultimately concerned with evaluators' responsibility for "doing the right thing" in terms of educational planning and programming and offers a perspective for moving in that direction. This higher order orientation has not always been present in evaluation theory or practice.

Naturalistic Evaluation*

One of the new methods and new concepts called for by Cooley and Lohnes' updating of Dewey's theory of valuation is that of naturalistic evaluation. An outgrowth of ecological psychology (Barker, 1965, 1968), naturalistic inquiry stands in contradistinction to the more formal models of evaluation previously discussed. Naturalistic evaluation has been referred to as an alternative to conventional evaluation methodology, breaking ties to both traditional forms of instrumentation and traditional methods of data analysis.

While many definitions of naturalistic inquiry have been proffered, Guba (1978) has suggested that naturalistic inquiry differs from other modes of evaluation by its relative position along two dimensions: (a) the degree to which the investigator manipulates conditions antecedent to the inquiry, and (b) the degree of constraint imposed on the behavior of subjects involved in the inquiry. Accordingly, naturalistic inquiry has been defined as

> ...any form of research that aims at discovery and verification through observation...(Willems and Rauch, (1969, p. 81)
>
> ...slice-of-life episodes documented through natural language representing as closely as possible how people feel, what they know, how they know it, and what their concerns, beliefs, perceptions and understandings are. (Wolf and Tymitz, 1976-1977)

*I am indebted to Guba (1978) for much of the material upon which this section is based. Readers are directed to his work for more on this topic.

> ...evaluation which attempts to arrive at naturalistic generalizations on the part of the audience; which is aimed at non-technical audiences like teachers or the public at large; which uses ordinary language; which is based on informal everyday reasoning; and which makes extensive use of arguments which attempt to establish the structure of reality. (House, 1977, p. 37)

In addition naturalistic studies have been identified by Sechrest as ones which

> (a) do not require the cooperation of the subject
> (b) do not permit the subject's awareness that he is being measured or treated in any special way, and
> (c) do not change the phenomenon being measured.
> (Willems and Rauch, 1969, p. 152)

In theory, a naturalistic study consists of a series of observations that are, alternately, directed at discovery and verification. This process supposedly leads to successive reorientations on the part of the investigator toward the phenomena being observed and to further discovery.

Unlike formal evaluation models, the naturalistic evaluator approaches data collection (observation) with a minimum of preconceived categories or notions of what will be seen, as though the behavioral phenomena were being observed for the first time. Any effort to manipulate any part of the program prior to observation or to constrain the behavior of those being observed would reduce the "naturalism" of the method. How data are tabulated and analyzed in a naturalistic study is left up to the investigator and no "best" method is identified, although it invariably includes some form of unstructured observation

followed by a piecing together of relationships, patterns or consistencies in the data which are used to further channel and focus subsequent observations. Data recording methods may include impressionistic accounts or ethnographic records* of the phenomenon observed. From these accounts more structured categories of behavior are derived, which then are expanded and verified through still further observation.

Naturalistic inquiry is appropriately considered by its proponents a tool, technique or method for viewing behavior and not exclusively a mode of evaluation. Thus, as a general methodology - or perhaps meta-methodology - its basic tenets would appear compatible with other forms or stages of evaluation which do not classify as experiments, i.e. where conditions are not prearranged and subject responses not constrained by the activities of the evaluator, e.g. goal-free evaluation. Naturalistic inquiry need not be considered an all-exclusive alternative to conventional models of evaluation when these other forms of inquiry do not unduly constrain the "naturalism" of the inquiry. Conducive to this line of reasoning is the idea that "naturalism" is always considered a matter of degree, making trade-offs and multiple approaches to evaluation possible. This view, however, receives little attention in the literature on naturalistic inquiry.

The extent and manner to which naturalistic inquiry has become inculcated in the present day thinking of evaluators is of considerable interest. The influence of naturalistic inquiry in this regard has been significant and represents what might be described as the underlying movement away from conventional

*A type of observational record associated with the field of anthropology in which behavior is recorded in relation to the context in which it occurs and is ascribed meaning only in relation to this context.

evaluation models and more formalistic definitions of evaluation, namely the measurement, congruency and applied research definitions. Oddly enough, it is a return of sorts to the visitation type definition of evaluation rejected by many evaluators a decade ago for being too subjective and impressionistic and represents in spirit, if not method, the value-oriented approach to evaluation. Value-oriented writers such as Dewey, Scriven and Apple would find solace in the fact that naturalistic inquiry, more than most other methodological perspectives, is likely to yield data unconstrained by preconceived notions about what the program is or is not supposed to do. This perspective seems congenial to the discovery of means-end relationships (Dewey), side effects and unanticipated program outcomes (Scriven) and fundamental issues which question the very rationale upon which a program is based (Apple).

While not embracing naturalistic inquiry directly, some evaluators have turned to this approach as a result of what are perceived to be serious limitations to conventional evaluation methods, namely: (a) that conventional models have been too restrictive in the types of data that can be observed and therefore may be insensitive to unique and unexpected program outcomes, (b) that conventional evaluation may at times actually contrive data by manipulating dimensions of a program which have no practical value in the real world and (c) that conventional modes of evaluation, particularly those ascribed to either the meansurement or congruency definitions of evaluation, may actually constrain through formal instrumentation the responses expected of subjects. In response to these limitations several evaluators have developed "alternative models" or approaches to evaluation which embody the elements of naturalistic inquiry. These models do not depend upon the arrangement of antecedent conditions or

constraint of subject response: hence, the basic conditions for naturalistic inquiry are met. These models, taken from Guba (1978), are reviewed briefly below. (For further explication of naturalistic inquiry see Guba, 1978, and Willems and Rauch, 1969.)

The Responsive Model. The first model with some relationship to naturalistic inquiry is the responsive model developed by Stake (1975, a,b). The responsive model focuses on important issues and concerns pertaining to a program. According to Stake, evaluation is responsive if it:

> orients more directly to program activities than to program intents; responds to audience requirements for information; and if the different value perspectives are referred to in reporting the success and failure of the program. (Stake, 1975, p. 14)

The primary purpose of responsive evaluation is to respond to audience requirements for information and to bring to the foreground different value perspectives that might be held by different audiences. Its methodology, like naturalistic inquiry itself, is nonconstraining. Stake describes it in the following terms:

> To do a responsive evaluation, the evaluator conceives of a plan of observations and negotiations. He arranges for various persons to observe the program and with their help prepares brief narratives, portrayals, product displays, graphs, etc. He finds out what is of value to the audiences and gathers expressions of worth from various individuals whose point of view differ. Of course, he

> checks the quality of his records: he gets program personnel to react to the accuracy of the portrayals; and audience members to react to the relevance of his findings. He does much of this informally - iterating and keeping a record of action and reaction. He chooses media accessible to his audiences to increase the likelihood and fidelity of communication. He might prepare a final written report, he might not - depending on what he and his clients have agreed on. (Guba, 1978, pp. 34-35)

These activities are carried out in a series of steps which may be described as (a) talking with clients, program staff and audiences, (b) identifying program scope, (c) providing an overview of program activities, (d) discovering purposes and concerns, (e) conceptualizing issues and problems, (f) identifying data needs relevant to the issues, (g) selecting observers and instruments (if any), (h) observing designated antecedents, transactions and outcomes, (i) thematizing or preparing portrayals in case studies, (j) winnowing, matching issues to audiences, (k) formating for audience use, and (l) assembling formal reports (if any).

The Judicial Model. A second evaluation model with some relationship to naturalistic inquiry is the judicial model. Developed by Wolf (1975), Owens (1973) and Levine (1974), the judicial model is patterned after the administrative hearing in a court of law. The purpose of the judicial model is to eliminate, inform and adjudicate issues related to the object or activity being evaluated. Advocates or counsels take opposite views with respect to an issue and argue as convincingly as possible their side of the issue. Jury and judge

hear testimony from "witnesses" and the presentation of facts regarding the issue, then offer their opinion as to the merit or worth of the program and their recommendations for improvement. Like the judicial process itself, this approach to evaluation assumes that "truth" is more likely to emerge in an adversary setting with two evaluators "pitted" against one another than in the case of a single evaluator using conventional evaluation models and data collection methods.

Generally, the following steps are employed in the judicial model:

(1) Issue generation. The issues are identified through "fact-finding interviews" with samples of the audiences involved, as in the case of the Stake responsive model.

(2) Issue selection. The purpose of this state is to delimit the number of issues and to prioritize them, so that they may be manageable in a hearing format.

(3) Preparation of formal arguments. Each counsel or advocate team prepares formal arguments related to the selected issues. Available evaluation or other data may be used (to be introduced as "exhibits" in the hearing stage), and additional evidence may be collected, particularly evidence in the form of depositions from witnesses. Additionally, selected witnesses may be asked to give testimony at the hearing itself.

(4) Pre-hearing discovery sessions. Each advocate team reviews the major arguments it intends to make and discloses the main features of its "evidence" for the other. Since the

hearing is not a "trial" in the conventional sense, but an effort to determine "truth" as precisely as possible, each side shares its findings with the other so that the hearing may be as comprehensive as possible. In addition, the advocate teams decide on ground rules, e.g., number of witnesses to be called and criteria for determining admissability of evidence.

(5) The hearing. Modeled on an actual courtroom process, the hearing involves an administrative officer and a "jury" or hearing panel. After hearing the evidence, the jury carries out whatever tasks the advocate teams previously agreed to assign to it, which usually involves at least the determination of findings (which may include judgments of worth) and the making of selected recommendations. (Guba, 1978, p. 36-37.)

The Transactional Model. A third evaluation model with some relationship to naturalistic inquiry is the transactional model described by Rippey (1973). This model supposedly differs from conventional models in that it deals directly with management conflicts and institutional change brought about by the implementation of a program, utilizing what its authors call "open systems theory." Transactional evaluation studies institutional disruptions brought about by the program and works to ameliorate these disruptions through strategies for conflict management.

Transactional evaluation has five phases:

(1) The initial phase. Pre-existing unrest or some other troublesome situation exists. A meeting is set up of

interested parties under the direction of a "neutral" evaluator working in a non-judgmental atmosphere.

(2) Instrumentation phase. During this phase, a "Transactional Evaluation Instrument" (TEI) is developed whose purpose is to provide the evaluator with insight into the perceptions and expectations of various interest groups. The instrument also provides a forum for the sharing of opinions among the groups. The TEI is developed and administered in group sessions, during which (a) the evaluator initially formulates issues on the basis of general expressions from the group, (b) participants are asked to re-express opinions about them, (c) the most representative and divergent of the written responses are carefully worded into items that can be rated on a scale from "strongly agree" to "strongly disagree," (d) the instrument is administered to the group, and (e) responses are examined.

(3) Program development. The program is redefined to reflect those goals and values on which the group can achieve some consensus.

(4) Program monitoring. Various groups agree to assume responsibility for implementing and monitoring the developed program.

(5) Recycling. As new conflicts emerge, the entire process is recycled to whatever phase is appropriate.

(Guba, 1978, p. 28; Talmadge, 1975)

The Connoisseurship Model. A fourth model with some relationship to naturalistic inquiry is the connoisseurship model developed by Eisner (1975). This approach views educational evaluation as a form of criticism. In Eisner's view, criticism depends upon connoisseurship - or the private act of appreciating and sensing the subtle qualities of an object or activity. "Critical guideposts" used to conduct the evaluation are essential elements of the connoisseurship approach. These guideposts represent the personal values and concepts formed from tradition, experience and theories about the standards for judging the object or activity. Guba (1978) characterizes connoisseurs as:

> persons with refined perceptual apparatus, knowledge of what to look for, and a backlog of previous relevant experience. They have the ability to recognize skills, form, and imagination and to perceive the intentions and leading conceptions underlying the entity being evaluated. In effect, because of these characteristics, the connoisseur is himself the evaluation instrument. Having made his judgments, he communicates the qualities that constitute the entity being evaluated, its significance, and the quality of experience engendered by interaction with it, often through the use of rich metaphors. (p. 39)

The Illumination Model. Perhaps most similar to naturalistic inquiry is the illumination model developed by Parlett and Hamilton (1977). This approach to evaluation relies heavily on open ended observations (but also questionnaires, interviews and tests) to continuously record ongoing events in order to (a) identify critical and nonobvious characteristics of a program, (b) the tacit

assumptions underlying it, (c) interpersonal relationships affecting it, and (d) complex realities surrounding the program. In the authors' words,

> illuminative evaluation, takes account of the wider contexts in which education programs function. Its primary concern is with description and interpretation rather than measurement and prediction. It stands unambiguously within the alternative methodological paradigm. The aims of illuminative evaluation are to study the innovatory program: how it operates; how it is influenced by the various school situations in which it is applied; what those directly concerned regard as its advantages and disadvantages; and how students' intellectual tasks and academic experiences are most affected. It aims to discover and document what it is like to be participating in the scheme, whether as teacher or pupil, and, in addition, to discern and discuss the innovation's most significant features, recurrent concomitants, and critical processes. In short, it seeks to address and to illuminate a complex array of questions.
>
> (Guba, 1978, p. 40)

Illuminative evaluation is carried out in three stages:

(1) initial observations for the purpose of familiarization with day-to-day reality of the setting(s), largely in the manner of social anthropologists or natural historians;

(2) more sustained and intensive inquiry into a number of common incidents, recurring trends, and issues frequently raised in discussion;

(3) efforts to seek general principles underlying the organization of the program, determine patterns of cause and effect within its operation, and place individual findings within a broader explanatory context. (Guba, 1978, p. 40)

Summary of Naturalistic Models

All five of the models presented qualify as naturalistic in that they adhere to the two primary conditions set forth by proponents of the naturalistic method: (a) they do not manipulate conditions antecedent to the inquiry and (b) they pose minimal constraints on the behavior of participants involved in the inquiry. While always a matter of degree, these five models meet these conditions to a greater extent than do most conventional approaches to evaluation.

However, it can also be noted that the five models are somewhat vague as to the precise manner in which observations are to be conducted and the data resulting from them converted into meaningful statements which serve some client group. Conspicuously lacking both in summary and original documents describing these models are descriptions of the processes by which responsive judicial, transactional, connoiseurship, and illuminatory accounts of behavioral phenomenon are gleaned of their most pregnant content and communicated to audiences who desire answers to specific questions, some of which may have been fashioned prior to program observation. If naturalistic methods are to enjoy widespread use, the criteria by which value and importance are bestowed upon the data may need further delineation within the context of each model. The absence of this delineation may result in what Kaplan (1974) has called "the dogma of immaculate perception." In explaining the importance of values in directing what the inquirer is looking for, Kaplan compares a value-free inquiry or one which limits itself to just describing what objectively happens

to the position of the esthetes at the turn of the century, who viewed art as a matter of pure form or decoration, "at the cost of making of it an idle song for an idle hour," with no significance for anyone but themselves.

We may also note that the concept of naturalistic inquiry was first introduced as an alternative methodology to present day conceptions of experimental design and not as an approach to serve the ends of evaluation. Although the authors of naturalistic models have done an exemplary job of making this relationship appealing, the match between naturalistic inquiry and evaluation may not be as great as it might at first seem. The decision-oriented context in which most evaluations occur are not always conducive to the hypothesis generating and theory building purposes for which naturalistic inquiry is best suited. Some audiences for evaluation studies may appreciate being confronted with "issues" and "concerns." But other audiences may not be so appreciative if specific questions requiring formal measurement and analysis are left unanswered simply because they require altering antecedent conditions or constraining subject responses. It is because of the diversity of what clients desire and expect of an evaluation that the word "supplementary" rather than "alternative" might be used to place naturalistic methods in its most appropriate framework.

Finally, it is important to note the perspective or mind set the naturalistic inquirer carries with him when studying behavior. This perspective, or Weltanschauung, (world-view) has been aptly captured by Louch (1966), who, in the context of describing the role of explanation in the study of human action, provides a good portrayal of the naturalistic inquirer and the commonality between the naturalistic and value-oriented approaches to evaluation. In the

words of Louch the world of the naturalistic inquirer is one in which:

> behavior cannot be explained by a methodology borrowed from the physical science." For him, "what is needed ...is not measurement, experiment, prediction, formal argument but appraisal, detailed description, reflection and rhetoric....Human action is a matter of appraising the rightness or appropriateness of what is attempted or achieved by men in each set of circumstances. Its affinities are with morality rather than with the causal or statistical accounts appropriate to the space-time framework of the physical sciences. Its methods are akin to the deliberations and judgments in the law rather than the hypotheses and experiments of physics.
> (Van Gigch, 1978, p. 220)

Systems-oriented Evaluation

While much of the evaluation literature of the past decade focused on distinctions between evaluation and research and the insensitivity of the latter to detecting the effects of innovative programs, conceptual models were being developed interconnecting the planning, development and evaluation process. These models, while not distinct from other approaches in their call to infuse the discipline of evaluation with a broader methodology than research, were distinct in their efforts to include within the domain of evaluation methodologies to actually improve the process by which programs were being planned and developed. To accomplish this purpose various systematic approaches to instructional development were introduced posing "front end" or predevelopment tasks for the evaluator,

unifying and integrating the previously separate processes of program planning, development and evaluation.

Kaufman (1972), in the first modern text dealing with educational planning from a systems perspective, defined system as:

> The sum total of parts working independently and working together to achieve required results or outcomes, based on needs. (p. 1)

and the systems approach as:

> A process by which needs are identified, problems selected, requirements for problem solution are identified, solutions are chosen from alternatives, methods, and means are obtained and implemented, results are evaluated, and required revisions to all or part of the system are made so that the needs are eliminated. (p. 2)

The particular systems approach articulated by Kaufman represents a type of logical problem solving for identifying and resolving educational problems. Central to this approach is the process of educational planning.

One example of the systems approach applied to planning and evaluation is the Interservice Procedures for Instructional Systems Development* (U.S. Army Training and Doctrine Command, 1975), a five-volume compendium on the "how to do it" aspects of instructional systems development. While developed for the military, this work represents a broad application of the systems approach to training useful in virtually any type of setting. The Interservice Procedures are divided into five separate and distinct phases to be carried out successively. These phases as described in the executive summary of the project are:

*Developed at the Center for Educational Technology, Florida State University.

Phase I, ANALYZE. This phase deals with procedures for defining what jobs are, breaking these into statements of tasks, and using numerical techniques to combine the best judgment of experienced professionals to select tasks for training. Phase I also presents processes for construction of job performance measures and the sharing of occupational and training information within and among client groups. It provides a rationale for deciding whether tasks should be trained in schools, on the job, or elsewhere, and also requires consideration of the interaction between training and job performance.

Phase II, DESIGN. This phase deals with the design aspects of the training program within selected settings. Design is considered in the architectural sense in which the form and specifications for training are laid down in careful detail.

Phase II reviews the considerations relating to entry behavior of two separate kinds: general ability, and prior experience. A rationale is presented for establishing requirements based on the realistic evaluation of both of these factors.

Phase III, DEVELOPMENT. This phase refers to the actual preparation of instruction. Determinations are made about how the students will be managed, the kinds of learning experiences they will have, the activities in which they will engage, and the form and content of the instructional delivery system. Techniques are presented for the careful review and adaptation of existing materials.

Procedures for the systematic design of instruction which can be delivered in a variety of media are also included. Phase III concludes with a procedure for testing and evaluating the instruction to insure that its performance meets expectations.

Phase IV, IMPLEMENTATION. This phase treats the necessary steps to implement the instruction according to the plan developed in Phase III. Two steps highlight Phase IV, that of training the staff in the procedures and problems unique to the specific instruction and actually bringing the instruction on-line and operating it. The Phase IV effort continues as long as there is a need for the instruction.

Phase V, CONTROL. This phase deals with procedures and techniques for maintaining instructional quality control standards and for providing data from internal and external sources upon which revision decisions can be based. Data collection, evaluation of the data, and decision making about the implications of the data represent the three principal functions described in Phase V. Emphasis is placed on the importance of determining whether the trainees are learning what was intended, and upon determining whether what they have learned is of benefit in carrying out post-training responsibilities.

These phases describe the functions necessary to analyze instructional needs; design, development, and implement instruction; and maintain quality control of instruction. Of primary importance is the sequential relationship of functions within and between phases, giving this model its systems perspective

In a similar manner, Dick and Carey (1978) integrate the processes of planning, development and evaluation into a ten step approach. These steps are: identifying instructional goals, conducting an instructional analysis, identifying entry behaviors and characteristics, writing performance objectives, developing criterion referenced tests, developing an instructional strategy, developing and selecting instruction, designing and conducting formative evaluation, revising instruction and conducting summative evaluation. Their procedure is described in some 200 pages and 10 chapters explicating each of these processes and integrating them into a single model. Other approaches have added still further to the language and conceptual repetiore of the systems approach, requiring the evaluator to conduct needs assessments, prepare program specifications, perform task and learner analyses and define human and material resources. More than simply terms and concepts these activities represent responsibilities which the systems-oriented evaluator is expected to perform.

Central to the systems approach is the blending of the humanistic and behavioralistic principles of psychology. The systems approach is considered humanistic in that it requires measurement of the needs of those the program is to serve. Through the conduct of needs assessments, the systems approach identifies discrepancies between "what is desired" and "what exists" and uses these discrepancies to provide direction for program development. Later, through program evaluation, the systems approach determines whether the desired state actually has been achieved. Needs assessments play a particularly central role in the systems approach by linking program design to extant needs for the purpose of improving program performance.

The systems approach derives its concepts and tools from a wide variety of disciplines including computer science, engineering, management science and economics. These tools are employed in the systems approach with the primary purpose of assuring that the program does what it is supposed to do. Accordingly, a systems approach to program development may specify rather elaborate procedures for assuring the accuracy and representativeness of the objectives upon which a program is to be based, for analyzing the characteristics of learners and the learning task and for monitoring the development process itself. These responsibilities have resulted in the blending into a single approach of concepts previously limited to either the field of instructional development or evaluation. This representation in a single approach of two previously distinct specialties has not been without its problems.

A major question around which some concern exists is whether the evaluator, especially formative evaluator, should be distinct from the developer or whether these roles represent responsibilities which can be fulfilled by the same individual working within the context of a systems approach. Some evaluators and developers warn that when role distinctions become unclear, as when an evaluator defines program requirements, conducts needs assessments and performs learner and task analyses, role distinctions become unclear and the program may suffer from what has come to be called co-option. This refers to the situation in which the evaluator is so emersed in the values, feelings and intents of the developer that evaluations are no longer an objective guide to program effectiveness. On the other hand, some evaluators and developers (Grobman, 1968; Butman and Fletcher, 1974) contend that development is so closely tied to evaluation that any separation of roles or functions is at

best an artificial distinction that may detract from rather than add to the development process. The popularity of "third party" or independent summative evaluations has dissipated to some extent the differences between these perspectives, when they are conducted in addition to formative assessments of program effectiveness and when the third party summative evaluator has not been provided knowledge of the outcome of the previous formative evaluations.

While little has been written about the role and function of the evaluator within the context of program planning, development and evaluation, it is not uncommon for a program to be planned and developed in such a way as to either encourage or preclude a certain kind of evaluation or that once the program has been developed the evaluator is forced to take a certain approach to evaluation regardless of its responsiveness to client needs. The systems approach argues, however, that the evaluator must serve critical functions early on in the development process to prevent just such an eventuality. Some of these "early on" evaluation activities are addressed below. (For the systems approach see also Banathy, 1968; Briggs, 1977; Davis, Alexander and Yelon, 1975 and Interservice Procedures for Instructional Systems Development, 1975.)

IV. Prospects for the 1980's: Implications of the Emerging Trends

The foregoing review of emerging trends has attempted to touch upon current evaluation theory and practice. This review has many implications for developers and evaluators of educational programs. These implications are generalizable to a variety of educational contexts be it elementary and secondary school, college, graduate school, inservice education or military training.

The purpose of this concluding section is to present several major implications of the emerging trends. These implications will be discussed generally and then illustrated with a specific advancement or change in evaluation practice which, in the opinion of the author, is likely to occur in the not-to-distant future. While the above trends represent an analysis of where evaluation is heading, the following implications represent signs or examples of the type of changes or advances which might be expected to result from these trends. These implications fall into the areas of systems approaches, naturalistic observation, needs assessment, policy assessment, and the role of the evaluator.

Implications for a systems approach

One implication of the emerging trends is that there is a need for a coherent, integrated approach to program planning, development and evaluation. Arguments have been presented that planning, development and evaluation can be seen as component parts of a unitary process, rather than conceptualized as separate and distinct activities. Program planning, especially, can be conducted with an eye toward program development (which it usually is) and program evaluation (which it usually is not). This implication can be reduced to a call for the application of a systems approach to instructional planning, development and evaluation.

Kaufman (1972) proposes that a systems approach to evaluation requires

the application of a variety of tools and techniques borrowed from the fields of computer science, cybernetics, engineering, management and operations research. These include simulations, operational gaming, the Program Evaluation and Review Technique (PERT), the Critical Path Method (CPM), the Delphi technique, and other systems analysis techniques. These tools are essentially modeling approaches to problem solving which fall under the rubric of systems analysis. While some of these modeling approaches have a distinct format and purpose, Kaufman (1972) advocates the use of graphic models for the general purpose of "displaying (or describing) a system and its components and subsystem relationships in a simple, 'at-a-glance' format" (p. 16).

Some recent developments in the field of general systems theory (Churchman, 1968, p. 155) have suggested that modeling as a means of studying a system may be useful for planning, developing and evaluating an educational program. Without guidelines on how systems modeling can be used to study educational programs, however, it is unlikely that the resulting models will be either communicative or generalizable across settings or applications. One implication for the not-to-distant future is the emergence of specific systems modeling techniques for decomposing or breaking down an instructional program (system) into its component parts prior to evaluation. Bloom et al. (1971) have already called for the use of such a technique, called a behavior by content matrix (or table of specifications), for understanding the nature of a developing program and guiding its evaluation. These authors suggest that a breakdown of the learning task "provides the specifications for formative evaluation and other procedures" (p. 17).

Ross (1977), Ross and Brackett (1976) and Ross and Schoman (1977), have suggested that a good system modeling technique should have certain specifiable properties. While referring to the development and portrayal of complex systems, the

properties posited by these authors are also applicable to the development and evaluation of educational programs. Their recommendations with some extensions and modifications to program evaluation are as follows:

1. Programs are best studied by building a model which expresses an indepth understanding of the program, sufficiently precise to serve as the basis for program development and evaluation.
2. Analysis of any program should be topdown (moving from general to specific outcomes), modular (take into consideration all component parts) and hierarchic (determine how the parts are tied together, i.e. structured).
3. Program activities should be represented by a diagram which shows program components, their interfaces, and their place in the hierarchic structure.
4. The model-building technique must represent behaviors the program is to produce, activities the program is to provide, and relationships among behaviors and activities.
5. All planning, design, development, and evaluation decisions should be in writing and available for open review to all team specialists.

These authors have developed a specific technique, the Structured Analysis and Design Technique (SADT)®, which meets these requirements and is applicable to the planning and evaluation of instructional programs.

Another implication of general systems theory for program evaluation is that a

program cannot be fully understood unless its relationship to the system in which it operates is known. Systems theory suggests that the behavioral changes often attributed to an instructional program are not due to the program alone but the interaction of the program with a milieu of variables comprising the environment of which it is a part. Simply put, systems theory suggests that more forces are at work than the program in effecting program outcomes and the more these other forces can be revealed through specific tools, such as program modeling, the greater the possibility of understanding and evaluating the program. While sometimes vague and illusive, instructional programs can be described in such a way (i.e. more precisely) that acknowledges the complex schema of person to person, person to environment, and environment to environment relationships in which they operate. To this end system analytic tools generally and system modeling techniques specifically can be useful in assisting evaluators to identify the contextual variables which moderate the effectiveness of instructional programs.

Implications for Naturalistic Inquiry

It is at this juncture that naturalistic inquiry and systems theory become reciprocally supporting concepts. Naturalistic inquiry, primarily procedures for observing behavior in naturally occurring settings, can provide a general tool with which the evaluator can identify and ultimately record the contextural factors which moderate a program's effectiveness. Our increasing awareness of the multidimensionality of the environment in which programs operate has led to the development of general methods by which this environment can be better understood. Naturalistic inquiry provides one such method for accomplishing this.

Recent advances in the behavioral and social sciences have made it increasingly difficult to understand important concepts or principles without viewing the

complex whole of which they are a part. The social and behavioral sciences have, in manner of speaking, run out of simple solutions. Or, more correctly, they have found simple solutions to old problems inadequate in light of recent discoveries and advancements which have all but nullified many "simple" views of instruction and behavior. Complexity is a fact of life and problem solving techniques which recognize this multidimensional environment seem particularly timely. This is why simplistic views of educational programs may no longer be credible and why naturalistic inquiry coupled with systems theory will be a useful tool for describing a program in terms of the larger system in which it operates. It is this extrospective - as opposed to introspective - view that brings systems theory and the aims of naturalistic inquiry together. Programs should not only be designed but also evaluated from a viewpoint which considers the effects larger systems (programs) have on smaller systems (programs). In the language of the systems approach, extrospective analyses trace program effects to contexts not ordinarily included in formal models of evaluation. These models commonly provide only for introspective analyses that trace program effects within the bounded context of the program under consideration.

Lastly, it is important, but unfortunate, to note that many programs are designed, operated and evaluated as though they were ends in themselves without considering that all programs are intended to satisfy the requirements of some larger system of which they are a part -- just as the objectives of a child's homework assignment is determined by the objectives of the unit of which it is a part and and the unit objectives determined by the subject matter of which it is a part and the subject matter determined by the objectives of the community who determine what is "best" for their children. Here is where the means-ends relationship which often goes unnoticed with conventional evaluation models could be

uncovered and evaluated with the systems approach. Systems theory and naturalistic inquiry provide a basis for identifying the means-end continuum, determining what program goals (ends) are not ends-in-themselves but means to still other ends, and whether the means actually justify the ends.

Consistent with the above notions are the following claims commonly ascribed to the systems approach (Van Gigch, p. 30):

(1) The system approach is indispensible in considering the relationships of a particular problem to its environmental conditions and in identifying the factors and variables that affect the situation.

(2) The systems approach brings out in the open inconsistencies of objectives when treating the various agents who play a part in the programs of the same system.

(3) The systems approach provides a useful framework in which the performance of the various systems, subsystems, and the whole system can be evaluated.

(4) The systems approach and its attendant methodology can be used to redesign the existing system and to compare and test the relative worth of alternative plans.

Implications for Needs Assessment

A third implication of the emerging trends derives from both the value-oriented and systems-oriented approaches to evaluation. This implication is that needs assessment is an evaluation activity that should be conducted at all stages of program development. Kaufman (1972, 1977) gives top priority to the needs assessment approach to evaluation. Indeed, in Kaufman's recent writing on systems (Kaufman, 1977), six types of needs studies were posited. The functions of these six types of needs studies are to: identify programs based upon needs (Alpha

type), determine solution requirements and identify solution alternatives (Beta type), select solution strategies from among alternatives (Gamma type), implement program (Delta type), determine performance effectiveness (Epsilon type), revise as required (Zeta type).

To Kaufman, a systems approach is a sequential series (Alpha-Zeta) of needs assessments -- a view which is consistent with a systems orientation. This series is presented in Table 5 along with some planning and evaluation tools suggested by Kaufman that are associated with each type.

Table 5

Planning and Evaluation Tools Available for Performing Each of the Functions of a System Approach*

Type of Needs Assessment	System Approach Function	Possible Planning Tools Associated with Each Function
1.0 Alpha Type	Identify problem based upon needs.	Needs assessment (Alpha type)
2.0 Beta Type	Determine solution requirements and identify solution alternatives.	System analysis, needs analysis, behavioral objectives, front-end analysis, performance analysis.
3.0 Gamma Type	Select solution strategies from among alternatives.	Systems analysis, Cost-effectiveness analysis, PPB(E)S, simulation, operations research/analysis, methods-means selection techniques, gaming.

Table 5
(continued)

4.0 Delta Type	Implement.	PERT, CPM, management by objective, management by exception.
5.0 Epsilon Type	Determine performance effectiveness.	Testing, assessment, auditing.
6.0 Zeta Type	Revise as required.	Discrepancy analysis. (Similar to a needs assessment.)

*from Kaufman, 1977

Another noteworthy aspect of Kaufman's (1977) approach is that it is hierarchical with respect to making faulty assumptions and achieving significant educational change. Kaufman states that the evaluator may start at any level of needs assessment but the further from the top (Alpha type) the level of entry, the lower the probability of actually achieving a meaningful change in educational practice and the greater the probability of making errors due to faulty assumptions. Value-oriented definitions of evaluation would require entry at the Alpha level (to identify needs and design the program). Decision-oriented definitions generally assume a lower entry level, sometimes as low as Zeta type needs assessment (to revise the program).

Needs assessment is usually conceptualized within a much narrower context than is reflected by Kaufman's Taxonomy. However, a wide variety of needs assessment type techniques and procedures are reported in the literature. These techniques include goal setting and goal rating procedures, strategies for assessing the current status of a program, discrepancy analysis, priority setting methods and various specialized techniques. (See a review by Witkin, 1977, describing these techniques and their advantages and disadvantages.)

Policy Assessment

A fourthimplication of the emerging trends is that policy assessments should be conducted early on in the planning process. Often confused with the concept of needs, policies represent a distinct area of inquiry which provide the data upon which needs studies are based. Policy studies come before needs studies and are used in deciding the type of needs study that should be conducted.

In a most general sense, policy assessment is the process by which one understands and anticipates the kinds of issues that are expected to result from alternative courses of action. These studies systematically examine the effects on society that may occur when an educational technology, program or product is introduced, extended or modified. Environmental impact statements now mandated of chemical processers, automobile producers, steel manufacturers, and airlines are examples of policy assessments. These assessments may be distinguished from needs studies by their attempt to:

1. Clarify goals
2. Identify assumptions behind goals
3. Define the consequences of goals, and
4. Portray alternative courses of action (goals).

These objectives entail the corollary tasks of identifying the parties who will be affected both directly and indirectly by the technology, program or product and describing the social, institutional, technological and economic factors which can change or be changed by the newly developed technology, program or product.

Contrary to the quantitative data which often earmark needs assessment,

policy assessment often results in a mix of hard and soft data. Kaufman's (1977) Alpha type needs assessment is, in part, a policy assessment when the inquiry is focused on desired goals. However, as Kaufman's taxonomy of needs assessments moves from a Beta to a Zeta type, the emphasis shifts from goals to means. As one moves down the list of Kaufman's five remaining types, the focus of the needs study changes to increasingly reflect means (Beta and Gamma types) or take means as givens (Delta, Epsilon and Zeta types). The fundamental difference between policy assessment and needs assessment is that the former focuses on the legitimacy of goals while the latter focuses on the legitimacy of means as illustrated by the following:

Topic of Policy Assessment	Topic of Needs Assessment
Should we contain the Soviets in East Africa?	How do we contain the Soviets in East Africa?
What alternatives are available to our energy crisis?	How do we move oil from Alaska's North slope?
Should we teach vocational education subjects to college bound students?	In what form should we teach vocational education to college bound students?
Should technological advancements, such as computer assisted instruction, be used in schools to perform direct instructional functions.	What technological advancements with direct instruction capability are suitable for the schools?

In each of the above, policy assessment focuses attention on the goal itself, the assumptions and consequences underlying it and/or alternative courses of action. Thus, policy assessment is more fundamental than needs assessment,

determining the area of inquiry for a needs assessment and the type of needs study most appropriate for a particular course of action. A needs assessment works most effectively in conjunction with a policy assessment which first must determine the appropriateness of a particular course of action.

The methodology of policy assessment is perhaps one of the fastest developing areas in the general field of evaluation. Methodological advancements in this area have spanned a broad array of qualitative and quantitative techniques often combining the two in unique ways. Coates (1976) has reported on a large list of these and has provided documentation as to their use. The following list focuses on those techniques from his list likely to be used by the evaluator in the not-to-distant future:

1. Trends extrapolation and futures-related techniques.
   Forecasting of the time of occurence of an event related to a particular goal.
   (Hencley and Yates, 1974)
2. Risk-benefit and fault-tree analyses.
   The codification of risks and assorted options under varying conditions of uncertainty.
   (National Academy of Engineering, 1971)
3. Delphi technique.
   Consensus forecasting among a panel of experts using cycles of information and feedback without face-to-face confrontation.
   (Linstone and Turoff, 1975)
4. Scenario, gaming and simulation.
   Mathematical and nonmathematical techniques for developing

4. complex statements of future conditions including psycho-dramatizations of current and existing states and simulations of future states.
   (Abt, 1970)
5. Cross impact analysis.
   A process whereby each individual prediction in a forecast is evaluated in relation to the probable truth or falsity of other predictions.
   (Monsanto, 1973)
6. Morphological analysis.
   A process by which all possible questions and answers pertaining to a certain problem are exhausted in a large question by answer matrix.
   (Zwicky, 1957)
7. Decision/relevance tree.
   A method for exhausting all possible options and alternatives with regard to a particular problem.
   (Gordon et. al., 1974; Gulick, 1979)
8. Judgment theory.
   A procedure which combines interrogation with statistical regression to assign weights to alternative courses of action.
   (Hammond and Summers, 1972)
9. Cost-benefit (input/output) studies.
   A class of econometric models using regression analyses for interrelating cost and productivity variables.
   (Leontief, 1966; see also Orlansky and String, 1978)

10. Structural and system dynamic modeling.

A group activity which applies logical reasoning to complex issues to determine interrelationships and networks among the elements of a system.

(Meadows, Meadows, Randers, and Behrens, 1972)

Perhaps the single most definable quality linking these techniques is the emphasis and importance they place on the participation of a broad mixture of experts and laymen who are capable of making judgments about the implications of policy. Most of these techniques represent highly democratic processes in which individual opinion is heavily weighted. In contradistinction to other forms of government where policy alternatives must be weighed against their compatability with an accepted ideology, these policy assessment techniques represent democratically oriented approaches to the generation of alternatives and definitions of consequences regardless of their relevance to the existing state of affairs. Thus, they represent relative approaches to problem solving where the criterion is unaffected by any absolute ideology but instead is defined by the best alternative available. This can be both an advantage and disadvantage in that (a) usually a large number of decision alternatives are generated by these methods (as in brainstorming), often making systematic data tabulation difficult, (b) usually small (but not necessarily insignificant) differences can exist among them, making evaluation of differences between alternatives difficult, and (c) no appeal to "higher" authority can be made to simplify the process, at least not initially. Only real-world constraints and implications that can be documented by logic or experience may be entered as acceptable data. On the other hand these policy techniques (a) provide for a maximum number of alternative courses of action to be discovered, many of which might not have been considered with less democratic methods, (b) represent

the best mix of opinions and viewpoints, often representing a consensus opinion reflecting the best features of individual viewpoints, and (c) lead to identification of alternatives for which there is high probability that practical strategies, solutions and methods actually exist.

The importance of the field of policy assessment over the coming decade will be directly linked to the extent to which new technological advancements and program development projects create new knowledge gaps and the extent to which the direct and unanticipated effects of a program proves to be as or more significant than the immediate or planned consequences of that program. The probable occurrence of both of the above should make policy assessment a major development in the field of evaluation in the next decade.

Role of the Evaluator

By now the reader is no doubt aware of the close and non-coincidental relationships which bind the concepts of systems modeling, naturalistic inquiry, and policy and needs assessment. The interrelationships among these four concepts are never more obvious than when their effects on the field of evaluation are seen through the role of the evaluator. We now conclude with a unifying theme which underpins these four concepts.

The emerging trends depicted in this paper as well as the implications above forecast an expanding role for the evaluator. This role will be shaped by an increasing tendency to define evaluation broadly and to include within this definition evaluation activities that are performed prior to program development. The systems approach to evaluation generally and the concepts of naturalistic observation, needs assessment and policy assessment specifically point to some of the ways the evaluator's role is expanding to include activities performed early on in the planning and development process.

The systems approach to evaluation confers on the evaluator a broad array of responsibilities heretofore divided among other specialists. While leaving the bulk of the planning and development work to these specialists, the systems approach places responsibility with the evaluator for many quasi-evaluation or concomitant activities which, while not a direct part of the planning and development process, hold potential for substantially improving the quality of program planning and development. These activities can so influence the design of an evaluation that their completion by the evaluator early on in the planning and development process may soon become a standard for good evaluation. An analysis of the systems approach can foreshadow some of these "front-end" activities the evaluator may soon be expected to perform.

The systems approach represents the integration of the planning, development and evaluation processes into a single coherent approach, thus implying an underlining thread by which these processes are linked. The responsibility for providing this link may increasingly fall upon the shoulders of the evaluator. The evaluator may be expected to perform this linking function with quasi-evaluation activities that accompany the process of evaluation but which are not themselves part of the act of determining the "merit or worth of a thing". Several such activities have already emerged such as policy assessments for determining the relative consequences of program objectives, needs assessments for determining the most appropriate methods, solutions and strategies for meeting program objectives, systems modeling techniques for clarifying and refining program design and naturalistic inquiry for determining the larger system in which the program must operate and, hence, means-ends relationships. While these activities can materially contribute to a unified approach to planning, development and evaluation they are not the only ones. The evaluator

can assume many other "front end" functions of a logical nature which help clarify and focus the work of the planner and developer. Determining the representativeness, accuracy and appropriateness of program objectives, the logicalness of intended relationships between program components and expected outcomes, the congruency of objectives with planned development activities, and the modeling of intended instructional activities and outcomes to depict hierarchical and sequential relationships among program objectives, are other activities implied by an integrated approach to planning, development and evaluation. These activities can be described as preformative activities, that is activities performed by the evaluator prior to program development. Because preformative evaluation flows from the policy and needs assessment process, it can be expected to influence all aspects of program planning, development and evaluation.

Traditionally, program planning, development and evaluation have been viewed as distinct roles or functions related in sequence but not substance. Formal training in evaluation has not always emphasized concepts of instructional design and development and vice versa. While the notion of formative evaluation has linked program development to evaluation, it has not related evaluation to program planning. Given the concepts of systems theory, naturalistic inquiry, and policy and needs assessment, a coherent, unified approach to planning, development and evaluation can emerge. The further development of these concepts and the linking of them to the planning, development and evaluation process is an important outcome for the field of evaluation in the decade ahead.

Postscript

The following observations were prepared after reviewing a small but representative sample of studies that have evaluated computer based instruction. Using the concepts and milestones discussed in earlier portions of this paper as an organizational framework, the evaluation issues, problems and concerns raised by these studies were noted and arranged in the following table. This table is intended as an argumentative "think piece" for discussing desired ways of evaluating computer based instruction.

Some Observations on the Evaluation of Computer Based Instruction (CBI) in Relation to Concepts and Milestones in the Field of Evaluation

Factors Contributing to the Growth of Evaluation	Their Relationship to the Evaluation of Computer Based Instruction
Educational Development and Societal Trends	For the studies reviewed, the evaluation of computer based instruction seemed to have followed many of the same trends that influenced the growth and development of the larger field of evaluation. Within this context the evaluation of CAI/CMI software development seemed most influenced by the principles of operationalism and behavioral objectives, moderately influenced by the ESEA Legislation of 1965 and least influenced by the school and teacher accountability movement. The use of models of software development stemming from the curriculum reform movement could also be noted, particularly the team authorship of instructional software and the formative evaluation and pilot testing of prototype materials.

Traditional Definitions of Evaluation

The evaluations of CBI that were reviewed were generally indistinguishable from applied research. The applied research character of these evaluations seemed to be regarded as a strength and not a factor limiting the generalizability or contextual validity of the conclusions that could be drawn from them. While these investigations were often called evaluations, they were, in essence, applied research studies indistinguishable from textbook definitions of experimental and quasi-experimental research. The statistical control of potentially confounding (interactive) variables, the terminal availability of data, the use of control groups and statistical decision rules (e.g., $p < .05$) were standard ingredients of these studies.

Models for Evaluation

Most CBI evaluations did not employ evaluation models. The applied research nature of these studies made the utilization of evaluation models difficult because many of these models fell outside the traditional definition of applied research. Given the applied research nature of these studies, the matching of any particular evaluation model to a particular CBI context was not possible. For example, only Provus' "cost-benefit" stage and Stufflebeam's "product" stage clearly matched the stated purposes of CBI evaluations, while such concepts as Stake's "logical contingency," Stufflebeam's "context evaluation" and Provus' "program definition" seemed too far removed from the conclusion oriented and hypothesis testing intent of these evaluations to be of use.

Most studies reviewed seemed to fit the PDK National Study Committee's definition of educational evaluation stated as:

> ...the process of delineating, obtaining, and providing useful information for judging decision alternatives.

Decision-oriented Evaluation

The purpose of many of these studies was to provide decision makers with information as to whether previously articulated program goals and objectives were being met. This purpose seemed to add a decidely summative emphasis to many evaluations resulting in considerably less emphasis on formative evaluation and program modification. Sometimes conclusions were used in an all or none manner: the program was either accepted or rejected, adopted or discontinued. But, other times the data were too ambiguous to point to such all-exclusive alternatives, and, hence, were ignored in making recommendations for program improvements.

Value-oriented Evaluation

CBI evaluations seemed to have paid little attention to the relationship between means and ends. Outcomes studied seldom went beyond end-of-instruction attitude and achievement indicators and seldom examined the extent to which the results of the computer based instruction served larger program or institutional goals. Goals and standards for a program were generally taken as givens and the program judged on the basis of how well it met these goals and standards. For example, better attitude and higher achievement were sometimes considered exclusive ends, even when the goal of the instruction might have also included the reduction of instructional time and successful on the job performance.

Naturalistic Inquiry

Naturalistic inquiry played little, if any, role in the CBI evaluations reviewed. If naturalistic inquiry is defined as "any form of research that aims at discovery and verification through observation..." (Willems and Rauch, 1969, p. 81) or "...slice of life episodes documented through natural language ..." (Wolf and Tymitz, 1976-1977), little in the manner in which these CBI evaluations were conducted would suggest such a theme. Thus, naturalistic inquiry as a series of observations that are alternately directed at the process of "discovery and verification" generally did not characterize CBI studies. Also, evaluators seldom approached the data collection with a minimum of preconceived categories or notions of what would be seen or as though the phenomena were being observed for the first time. Thus, data were tabulated and analyzed in traditional ways with traditional statistical techniques, allowing little to be discovered other than that which was expected at the onset of the study.

Systems Approach

If the systems approach is viewed as a coherent, integrated approach to planning, development and evaluation, then few CBI projects evidenced this concept. To the contrary, planning, development and evaluation were generally compartmentalized with the evaluator arriving at the end of the development task to fill an applied research or summative role. Activities such as defining needs, conducting task analyses, identifying entry behaviors and writing performance objectives, while sometimes implicitly carried out by program developers and managers, generally were not considered part of the process of evaluation and seldom were conducted in a systematic manner. Other characteristics of the systems approach, such as its capacity to deal simultaneously with multiple dimensions of the instructional environment, its reliance on program modeling to describe this environment, and the conceptualization of parts within wholes (programs within larger programs), also were not in evidence. CBI evaluations seldom considered the effects of the system in which the program operated or traced program effects to other than the immediate stimuli under consideration.

Needs and Policy Assessment

Absent from most CBI evaluations was a determination of the needs and policies upon which the design for a particular program was based. This led to the failure of many studies to actually document whether the program's performance met some empirically determined need. Also, the policies of the agencies or institutions in which CBI programs operated, while perhaps implicitly known, were seldom empirically determined. These policies, if documented, might have in some instances resulted in the selection of different control treatments with which the computer based instruction was to be compared by ruling out the feasibility of some forms of instruction for technical, administrative or economic reasons. Most importantly, needs and policy studies were not used to pinpoint the precise nature of the problem which justified development of the computer based instruction in the first place. The reason why CBI was considered a more reasonable alternative than some other mode of instruction was seldom made clear and this, in turn, focused evaluations on CBI generally rather than on those unique aspects of the medium which could account for its superiority over an alternative treatment.

Role of the Evaluator

The role of the evaluator was essentially one of an applied researcher. Typically the evaluator's role consisted of instrument development, data collection, statistical analysis and report writing. Seldom did it allow for the determination of needs and policies relevant to the design of a program or for determining the ultimate end to which CBI was to be the means. Generally, CBI evaluations did not define evaluation broadly to include preformative activities which could determine the needs and justification upon which a program might be based. This role for the evaluator was most consistent with the decision-oriented definition of evaluation wherein the goals and objectives for a program are taken as givens. This role was less consistent with the value-oriented definition of evaluation wherein the work of the evaluator includes determination of the merit, worth or value of the goals and objectives themselves.

References

Abt, C. Serious games. New York: Viking Press, 1970.

Alderman, D. L. Evaluation of the Ticcit computer-assisted instructional system in the community college. Princeton, N.J.: Educational Testing Service, 1978.

Apple, M. W. The process and ideology of valuing in educational settings. In M. Apple, M. Subkoviak, and H. Lufler, Jr. (eds.) Educational evaluation: Analysis and responsibility. Berkeley, CA.: McCutchan, 1974.

Banathy, B. Instructional systems. Belmont, CA.: Fearon Publishers, 1968.

Barker, R. G. Ecological psychology. Stanford, CA.: Stanford University Press, 1968.

Barker, R. G. Explorations in ecological psychology. American Psychologist, 1965, 20.

Becker, H. Whose side are we on? In G. Riley (ed.) Values, objectivity and the social sciences. Reading, Mass.: Addison-Wesley, 1974.

Belliott, F. Design for the Tennessee assessment and evaluation of Title III, ESEA. Nashville: State of Tennessee Department of Education, 1969.

Bloom, B. (ed.) Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain. N.Y.: McKay, 1956.

Borich, G. (ed.) Evaluating educational programs and products. Englewood Cliffs, N.J.: Educational Technology Publications, 1974.

Borich, G. D. The appraisal of teaching: concepts and process. Reading, Mass.: Addison-Wesley, 1977.

Borich, G. D. Program evaluation: New concepts, new methods. Focus on exceptional children, 1977, 9 (3), 1-16.

Bridgman, P. The logic of modern physics. New York: MacMillan, 1927.

Briggs, L. (ed.) Instructional design. Englewood Cliffs, N.J.: Educational Technology Publications, 1977.

Butman, J. W. and J. L. Fletcher. The role of the evaluator and developer in educational research and development. In G. Borich (ed.) Evaluating educational programs and products. Englewood Cliffs, N.J.: Educational Technology Publications, 1974.

Campbell, D. and J. Stanley. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Churchman, C. W. A challenge to reason. New York: McGraw-Hill, 1968.

Coates, J. Technology assessment - a tool kit. Chemtech, June, 1976, 372-383.

Cooley, W. and P. Lohnes. Evaluation research in education. New York: Irvington Publishers, 1976.

Davis, R., L. Alexander and S. Yelon. Learning systems design. New York: McGraw-Hill, 1975.

Dewey, J. Human nature and conduct. New York: Henry Holt, 1922.

Dewey, J. Theory of valuation. International Encyclopedia of Unified Science, 1939, 2 (4), 1-67.

Dick, W. and L. Carey. The systematic design of instruction. Glenview, Ill.: Scott, Foresman and Co., 1978.

Dykstra, R. Summary of the second-grade phase of the cooperative research program in primary reading instruction. Reading Research Quarterly, 1968, 4 (1), 49-70.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, N. J.: Prentice-Hall, 1965.

Eisner, E. Instructional and expressive objectives: Their formulation and use in curriculum. In AERA monograph series on curriculum evaluation, No. 3: Instructional objectives. Chicago: Rand McNally, 1969.

Eisner, E. W. The perceptive eye: Toward the reformation of educational evaluation. Occasional papers of the Stanford Evaluation Consortium, Stanford, CA.: Stanford University, 1975 (mimeo).

Emrick, J., P. H. Sorenson and M. Stern. Interim evaluation of the national follow through program 1969-1971: A technical report. Menlo Park, CA.: Stanford Research Institute, 1973.

Furst, E. J. Constructing evaluation instruments. New York: David McKay, 1964.

Gordon, T. et al. A technology assessment of geothermal energy, R 164-46-11, The futures group, Clastonbury, Conn., 1974.

Grobman, H. Evaluation activities of curriculum projects: A starting point. In AERA monograph series on curriculum evaluation, No. 2: Evaluation activities of curriculum projects. Chicago: Rand McNally, 1968.

Guba, E. Toward a methodology of naturalistic inquiry in educational evaluation. Los Angeles: Center for the Study of Evaluation, University of California, 1978.

Guba, E. The failure of educational evaluation. Educational Technology, May, 1969, pp. 29-38.

Gulick, R. M. Decision analysis as a learning strategy. In O'Neil, H. and Spielberger, C. (eds.) Learning strategies: Issues and procedures. New York: Academic Press, 1979.

Hammond, R. Evaluation at the local level. In B. Worthen and J. Sanders, Educational evaluation: Theory and practice. Worthington, Ohio: Charles A. Jones, 1973.

Hammond, K. and D. Summers. Cognitive control. Psychological Review, 79, 1972, 58-67.

Handy, R. Value theory and the behavioral sciences. Springfield, Ill.: Charles C. Thomas, 1969.

Handy, R. The measurement of values. St. Louis, Mo.: Warren H. Green, 1970.

Handy, R. and P. Kurtz. A current appraisal of the behavioral sciences. Great Barrington, Mass.: Behavioral Research Council, 1964.

Hemphill, J. The relationships between research and evaluation studies. In R. Tyler (ed.) Educational evaluation: New Roles, new means, the sixty-eighth yearbook of the National Society of the Study of Education. Chicago: University of Chicago, 1961.

Hencley, S. and J. Yates. Futurism in education. Berkeley, McCutchan, 1974.

Hoban, C. Educational technology and human values. Audiovisual Communications Review, 1977, 25 (1), 221-242.

House, E. R. The logic of evaluative argument. CSE Monograph Series in Evaluation, No. 7. Los Angeles: Center for the Study of Evaluation, University of California, 1977.

Interservice procedures for instructional system development (IPISP). Ft. Monroe, VA: U.S. Army Training and Doctrine Command (TRADOC), Pamphlet 350-30, August, 1975.

Jemelka, R. and G. Borich. Traditional and emerging definitions of educational evaluation. Evaluation Quarterly, in press.

Kac, M. Some mathematical models in science. Science, 166, No. 3906695 (1969).

Kaplan, A. Values in inquiry. In G. Riley (ed.) Values, objectivity, and the social sciences. Reading, Mass.: Addison-Wesley, 1974.

Kaufman, R. Educational system planning. Englewood Cliffs, N. J.: Prentice-Hall, 1972.

Kaufman, R. Identifying and solving problems: A systems approach. LaJolla, CA.: University Associates Publishers, 1976.

Kaufman, R. A possible taxonomy of needs assessments. Educational Technology, 1977, 17 (11), 60-64.

Krathwohl, D., B. Bloom, and B. Masia. Taxonomy of educational objectives: The classification of educational goals. Handbook 2. Affective domain. New York: McKay, 1974.

Kuhn, T. S. The structure of scientific revolutions. Chicago: University of Chicago Press, 1970.

Leontief, W. Input-output economics. N. Y.: Oxford University Press, 1966.

Levine, M. Scientific method and the adversary model. American Psychologist, 1974, 29.

Linstone, H. and M. Turoff, (eds.) The delphi method: Techniques and applications. Reading, Mass.: Addison-Wesley, 1975.

Louch, A. R. Explanation and human action. Berkeley and Los Angeles, University of California Press, 1966.

Mager, R. F. Preparing objectives for programmed instruction. San Francisco: Fearon, 1962.

Meadows, D. H., D. L. Meadows, J. Randers and W. Behrens. The limits to growth. N. Y.: Universe Books, 1972.

Metfessel, N. W. and W. B. Michael. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. Educational and Psychological Measurement, 1967, 27, 931-943.

Monsanto Co. Environmental impact of nitrile barrier containers: LOPAC; A case study. St. Louis, Mo.: Symp. Proc., Monsanto Co., July 19, 1973.

Murphy, R. T. and Appel, L. R. Evaluation of the Plato IV computer-based education system in the community college. Princeton, N.J.: Educational Testing Service, 1977.

National Academy of Engineering. Perspectives on benefit-risk decision-making. Washington, D. C.: Rep. Colloquium, Committee on public engineering policy, National Academy of Engineering, April 26-27, 1971.

Orlansky, I. and J. String. Cost effectiveness of computer based instruction in military training. Arlington, VA.: Institute for Defense Analysis, Science and Technology Division, 1978.

Owens, T. R. Educational evaluation by adversary proceedings. In E. R. House (ed.), School evaluation: The politics and process. Berkeley, CA.: McCutchan Publishing Corporation, 1973.

Parlett, M., and D. Hamilton. Evaluation as illumination: A new approach to the study of innovatory programs. In G. V. Glass (ed.), Evaluation studies review manual (Volume 1). Beverly Hills, CA.: Sage Publications, 1976.

Popham, W. Objectives and instruction. In AERA monograph series on curriculum evaluation, No. 3: Instructional objectives. Chicago: Rand McNally, 1969.

Poynor, H. (ed.) Problems and potentials of educational evaluation. Austin, Texas: Southwest Educational Development Laboratory, 1974.

Provus, M. Discrepancy evaluation. Berkeley, CA.: McCutchan, 1971.

Rippey, R. M. Studies in transactional evaluation. Berkeley, CA.: McCutchan Publishing Corporation, 1973.

Ross, D. Structured analysis (SA): A language for communicating ideas. IEEE Transactions on Software Engineering, 1977, SE-3 (1), 16-34.

Ross, D. and J. Brackett. An approach to structured analysis. Computer Decisions, 1976 (Sept.), 40-44.

Ross, D. and K. Schoman. Structured analysis for requirements definition. IEEE transactions on software engineering, 1977, SE-3 (1), 6-15.

Scriven, M. The methodology of evaluation. In AERA monograph series on curriculum evaluation, No. 1: Perspective of curriculum evaluation. Chicago: Rand McNally, 1967.

Scriven, M. Goal free evaluation. In E. House (ed.) School evaluation, Berkeley, CA.: McCutchan, 1973.

Scriven, M. The concept of evaluation. In M. Apple, H. Subkoviak, and H. Lufler, Jr. (eds.), Educational evaluation: Analysis and responsibility. Berkeley, CA.: McCutchan, 1974.

Smith, E. and R. Tyler. Appraising and recording student progress. (Adventure in American Education Ser., Vol. 3) New York: Harper, 1942.

Stake, R. E. Evaluating the arts in education. Columbus, Ohio: Charles E. Merrill Publishing Company, 1975a.

Stake, R. E. Program evaluation, particularly responsive evaluation (occasional paper no. 5) The Evaluation Center, Western Michigan University, 1975b.

Stake, R. E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

Stufflebeam, D. L., W. J. Foley, W. J. Gephart, E. G. Guba, R. L. Hammond, H. D. Merriman, and M. M. Provus. Educational evaluation and decision making. Itasca, Ill.: Peacock, 1971.

Talmadge, H. Evaluation of local school community problems: A transactional evaluation approach. Journal of Research and Development in Education, 1975, 8.

Thorndike, R. L. and E. Hagen. Measurement and evaluation in psychology and education. New York: Wiley, 1969 (2nd ed.).

Tyler, R. W. Basic principles of curriculum and instruction. Chicago: University of Chicago Press, 1950.

Van Gigch, J. P. Applied general systems theory. New York: Harper and Row, 1978.

Willems, E. P. and H. L. Rauch. Naturalistic viewpoints in psychological research. New York, N. Y.: Holt, Rinehart and Winston, 1969.

Wolf, R. L. Trial by jury: A new evaluation method. Phi Delta Kappan, 1975, 57.

Wolf, R. L. and B. Tymitz. Ethnography and reading: Matching inquiry mode to process. Reading Research Quarterly, 1976-77, 12.

Worthen, B. and J. Sanders. Educational evaluation: Theory and practice. Worthington, Ohio: Charles Jones, 1973.

Zwicky, F. Morphological astronomy. New York: Springer-Verlag, 1957.